지리학석사 학위논문

Geographical Distribution and Regional Disparity of Suicide Risks in Korea, 2008-2012

국내 자살위험의 지리적 분포와 지역적 불균등에 관한 연구, 2008-2012

2015 년 2 월

서울대학교 대학원

지리학과

엄 영 섭

Geographical Distribution and Regional Disparity of Suicide Risks in Korea, 2008-2012

지도 교수 박 기 호

이 논문을 지리학석사 학위논문으로 제출함
2015 년 2 월

서울대학교 대학원
지리학과
엄 영 섭

엄영섭의 지리학석사 학위논문을 인준함
2014 년 12 월

위 원 장 _______________________ (인)

부위원장 _______________________ (인)

위 원 _______________________ (인)

Abstract

The suicide problem has been grave national public health issue. This study aims to explore geographical distribution of suicide risks in regional scale and to measure the regional disparity.

The study object is complete suicide, so the national mortality data, cause of death statistics, was used for this study. From the explorative analysis, significant differences of suicide rates among different sex and age strata were found. The associations between categorical socio-demographic variables and violent suicide means selection, holiday effect, and copycat effect were identified with log-linear model.

Spatio-temporally comparable suicide risks and statistically robust relative risks were estimated by direct standardization and Bayesian estimation, respectively. Various visualizing techniques, such as building legend scheme, statistical significance mapping, and cartogram, were applied for mapping the risks and the results were produced with interactive maps, utilizing the interface of Google Maps and Google Earth. In metropolitan areas, “bull’s eye” pattern was found. On the national scale, low suicide risks in southern areas were clearly distinguished with high risks in northern parts. Taking into account statistical significance, the regions with high relative risks were mainly distributed in Gangwon, Chungnam, and Chungbuk.

The regional disparity of suicide risks was measured with spatial dissimilarity index, which integrates two divergent dimensions: aspatial unevenness and spatial clustering. First, the characteristics of the distribution of suicide risks were measured in each dimension. Regional unevenness was measured with aspatial inequality indices and global/local spatial autocorrelation was explored with spatial cluster analysis. The regional inequality decreased from 2008 to 2010, and increased consistently after 2010. Moreover, strong spatial autocorrelation was observed throughout the study period. Based on these observations, the global spatial dissimilarity index was computed with Gaussian kernel. It did not make any change in the pattern of temporal variation derived with aspatial measures due to the stable spatio−temporal distribution of suicide risks. However, when the administrative units are ignored with kernel smoothing, temporal pattern changed, which implies the possibility of MAUP. Lastly, cumulative frequency legend and 3D mapping was applied to suicide atlas to reflect the information on the computed regional disparity of suicide risks.

Keyword : suicide risk, relative−risk estimation, suicide atlas, regional disparity, spatial cluster, spatial dissimilarity

Student Number : 2009−20152

Table of Contents

List of Tables

List of Figures

Ⅰ. Introduction

1.1. Background

It is estimated that almost one million people of the world commit suicide every year (World Health Organization 2012). Suicide is a public health problem worldwide, most notably in South Korea (hereafter, Korea). After the suicide rates have shown upward trend since early 1990s, suicide problem has been recognized as a serious public health issue in Korea. Suicide has been the fourth leading cause for 6 consecutive years since 2007 (Korea National Statistics, 2008, 2009, 2010, 2011, 2012, 2013). Especially, for teenagers and young adults, suicide has been the top leading cause of death (Table 1). Moreover, intentional self-harm has been the most frequent cause of death in middle aged people except malignant neoplasm. Considering the fact that suicide is the only selectable cause of death, which is completed by intention, it is a severe public health problem.

Table 1. The top three causes of death by age (2007-2012)

Age	2007			2008			2009			2010			2011			2012		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1-9	T	M	C	T	M	C	T	M	C	T	M	C	M	T	C	M	T	C
10-19	T	I	M	T	I	M	I	T	M	I	T	M	I	T	M	I	T	M
20-29	I	T	M	I	T	M	I	T	M	I	T	M	I	T	M	I	T	M
30-39	I	M	T	I	M	T	I	M	T	I	M	T	I	M	T	I	M	T
40-49	M	I	L	M	I	L	M	I	L	M	I	L	M	I	L	M	I	L
50-59	M	B	L	M	L	B	M	I	H	M	I	L	M	I	H	M	I	H
60-69	M	B	H	M	B	H	M	B	H	M	B	H	M	B	H	M	H	B
70-79	M	B	H	M	B	H	M	B	H	M	B	H	M	B	H	M	B	H
80+	M	B	H	M	B	H	M	B	H	M	B	H	M	H	B	M	H	B

* Source: Korea National Statistical Office, Cause of death statistics (2008-2013)
* Note: T: Transport Accident / M: Malignant neoplasm (Cancer) / C: Congenital malformations, deformations and chromosomal abnormalities / L: Diseases of liver / B: Cerebrovascular diseases / H: Heart Diseases / I: Intentional self-harm (Suicide; Bold)

With the gravity of the situation, government established comprehensive countermeasures in 2004 and 2009, for preventing mortality of suicide. Also, 'Act for the prevention of suicide and the creation of culture of respect for life' was enacted in 2011 and has been enforced from 2012. Even though there have been national effort to reduce suicide mortality, it is still one of the most serious public health problem in Korea.

Many studies have been performed in Korea to address this national problem. Previous studies primarily put their interests on the temporal variations of suicide and related risk factors, such as economic crisis (Kim et al. 2004, Chang et al. 2009), daily temperature (Kim, Kim and Kim 2011b) and divorce (Yip et al. 2012); on the copycat suicides driven by undiscerning broadcasting of media (Ji et al. 2014); on the suicide methods (Chen, Park and Lu 2009, Im et al. 2011, Ahn et al. 2012). In addition, most of previous studies performed their analysis at the national scale.

One important fact is that geographical pattern of suicide has been overlooked so far. It is needed to analyze suicide from the geographical perspective, because geographical distributions of suicide and the spatial patterns may provide the alternative approach to address the national health problem.

1.2. Objectives

In this social and scientific background, this study aims to analyze suicide data from geographic perspectives and provide information on the domestic suicide in regional scale. Specifically, this study has two objectives.

One of the purpose of this study is to manifest geographical distribution of suicide in analytical ways. For this objective, map can be used as the most effective tool to represent and explore current state of health issues in spatial context. Moreover, it can visualize unequal regional distribution of suicide, which cannot be accomplished from the tables of mortality rates (Dorling and Barford 2007). Established procedures utilized in disease mapping and suicide studies will be reviewed and followed, and it would be possible to consider additional mapping techniques which can be used to convey information in more efficient way. The visualization of suicide risks would help to recognize spatial pattern of health risks and contribute to decision−making for prevention.

Moreover, this study aims to identify how unequally the regional suicide risks are distributed. Health inequality is one of the key issues in medical and health geography. The quantification of regional disparity of suicide enables to summarize the uneven geographical distribution and identify temporal changes of inequality during the study period. Furthermore, the numerical summaries can reduce misleading of map readers and guarantee objectivity on the explanation of the spatial pattern, providing implications on effective policy−making.

1.3. Organization of Chapters

To achieve the objectives, this study was organized as Figure 1. Previous studies on disease mapping in medical geography and suicidal studies, and on measurements of regional inequality will be reviewed. From the literature review, both established and alternative methodologies which would be used in this study will be found. Then, detailed methodologies, required to achieve research objectives of this study, will be articulated. Analyses will be conducted with application of the methodologies, and the results will be described. Lastly, the findings, limitations, and research ideas for further analyses will be discussed.

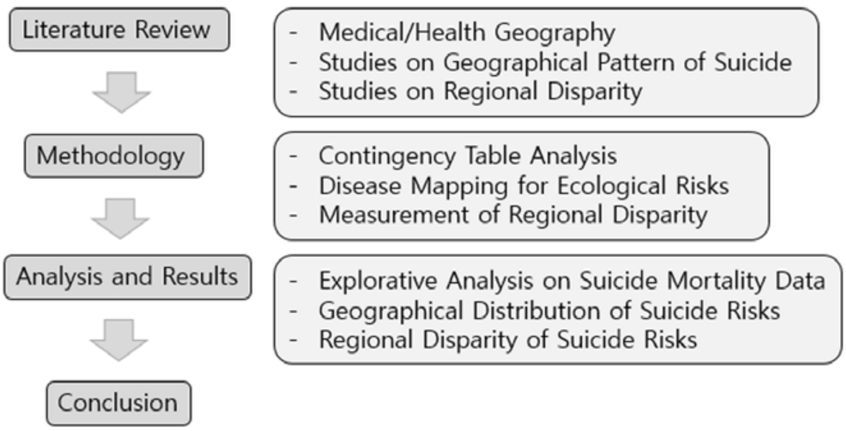


Figure 1. Organization of research

Ⅱ. Literature Review

2.1. Medical Geography and Disease Mapping

Public health is inherently related to geography. Epidemiological studies have been implemented from geographic perspective to investigate public health, and be referred to as medical geography. Because medical geography does not treat topics which is actually "medical", it is better to be acknowledged and labeled "epidemiological geography" (Mayer 2010). It have been developed in two distinct ways: disease ecology and health−care service. Disease ecology have been focused on the ill−health in populations like epidemiology, thus also being called as spatial epidemiology. Geographical coverage of health−care and access to the service have been analyzed as the second subgroup of medical geography. Health geography was introduced as the third approach, requesting reform of medical geography in 1990s (Kearns 1993). With the advent of health geography, there have been debate on the construction of sub−disciplines, while they have expanded their scopes and scales (Brown, McLafferty and Moon 2009). Health geography have broadened topics in critical, cultural and social geography (Dorn, Keirns and Del Casino Jr 2009), focusing mainly on health, place, and welfare (Kearns and Collins 2010). While health geography primarily takes qualitative methods and humanistic approach, medical geography relies on quantitative methods and spatial analysis. Mayer (2010) acknowledged the contribution of health geography, but also stressed the advantageousness of disease ecology combined with sophisticated spatial modelling and mapping techniques.

Disease mapping has been commonly used in health analysis because it gives intuition on the geographical pattern of health outcome of interest. John Snow's map of cholera (Snow 1855) is the most known case of disease mapping. First national atlas, which aimed to represent geographical patterns of specific cause of deaths, was made in England in 1964 (Howe 1964). This atlas used

Standardized Mortality Ratio (SMR) and built legend with color arrangement of black–white and graphical pattern, such as horizontal and diagonal pattern. SMR have been widely used from the time as a common mapping technique in disease mapping. With the success of atlas in England, enhanced access to large–scale data, and the development of computing system, national disease atlas started to be made in U.S. from 1975. Many disease atlas have been made from late 1990s, with the development of methodologies of disease mapping (Pickle 2009).

Recently, disease mapping have focused on exploring effective methodologies to augment the public awareness on the health issue. It is crucial that disease map should be easily accessed and available rather than be hidden from public audiences (Beyer, Tiwari and Rushton 2012). Thus, the importance of web–based mapping and publishing suicide data has been acknowledged. The developments of computing capability and web–based visualizing technique enabled practitioners to analyze the mortality data more efficiently and share the result more widely on the Internet. Moreover, web–based map is much more cost–effective in publishing (Peterson, 2003: 8), while it takes a lot of time and costs to publish national atlas in physical form. In addition, web–based map can easily modified even after publishing, contrary to the physical formatted maps. Above all, web–mapping is advantageous in the way that it is interactive. User can easily magnify, minimize, and pan the map screen, and identify useful information locating their mouse point to the region of interest. Also, the information can be spread widely even within a short time period, and shared spontaneously. With interoperability and the power of influence, web atlas is expected to enhance the comprehension of the public on the regional information.

2.2. Geographical Pattern of Suicide

It has been debated whether suicide is personal matter or a problem influenced by environment. Originally, suicide has long been regarded as individual problem. However, after Durkheim (1952) set forth a counterargument to debunk the established theory, it is now widely acknowledged that health behavior of individual is strongly affected by environment. According to Durkheim's theory, even though individual mental disease or race could affect suicide, the extent to which non−social factors lead to suicide is not substantial; suicide is rather mostly influenced by social causes, such as poverty and social isolation (Durkheim 1952). The completion of suicide depends on each individual's choice. However, there will be certain environmental factors which drive them to make the tragic decision. In advance to revealing contextual effect of suicide, it is important to analyze suicide in macro level, including identification of high−risk areas. To explore and investigate the geographical distribution of suicide, Durkheim (1952) mapped suicide rates, revealing that suicides were concentrated in the center of Europe, with minimum rates in the South and North of Europe. Also, he pointed out utilizing maps, comparing the map of suicide rate with those of putative variables, such as alcoholism to unveil spatial relationship between them.

Studies on mapping and investigating spatial heterogeneity of suicide have been mainly performed in England and Wales (Saunderson and Langford 1996, Middleton, Sterne and Gunnell 2008, Gunnell et al. 2012). Recently, the attempts to reveal spatial pattern of suicide have been made in other countries, such as Taiwan (Chang et al. 2011), Brazil (Bando et al. 2012a, Bando et al. 2012b), Australia (Qi, Tong and Hu 2010, Cheung et al. 2012) and Hungary (Balint et al. 2014).

In methodological view, to visualize the geographical distribution of suicide, SMR has been widely used to calculate relative risks (RRs) and visualize them (Saunderson and Langford 1996, Middleton et al. 2008, Chang et al. 2011, Cheung et al. 2012, Gunnell et al. 2012).

Moreover, Bayesian smoothing techniques have been applied to derive RR which is statistically robust: empirical Bayes (Saunderson and Langford 1996) and hierarchical Bayesian method (Middleton et al. 2008, Chang et al. 2011, Cheung et al. 2012, Gunnell et al. 2012). However, some studies argued that suicide rates should be standardized in direct method(Pickle 2009), and mapped Directly Standardized Rates (DSR) of suicide (Rezaeian et al. 2007, Qi et al. 2010).

It is often debated whether suicide rates should be standardized directly or indirectly. Beyer et al. (2012) suggested to consider both method in standardization as an essential property of disease mapping, because both are useful with their own purpose. The direct method is used when the purpose of the mapping is comparison of suicide rates across different areas, while indirect standardization is useful to represent health burden in each area. A study implemented in Australia used both standardization method to examine the geographical difference from place to place (Qi et al. 2012).

2.3. Measurement of Regional Inequality

2.3.1. Inequality Measures in Health Studies

Health inequality is the varying health outcome between individuals and groups. Previous studies have primarily focused on the health inequality in individual level, or on the inequality among socioeconomic groups (Wagstaff, Paci and van Doorslaer 1991, Lorant et al. 2005, Braveman 2006, Holstein et al. 2009). The regional health inequality is an application of inequality measures on the “geo-referenced areal units” (Rey and Smith 2013). Theil index was pointed out as a useful tool for assessing relative regional inequality and the inequality within non-ordered group, especially with decomposition of total inequality into between-group inequality and within-group inequality (World Health Organization 2013). Moreover, Fang et al. (2010) developed a composite index for health and evaluated regional inequality of health in China with Lorenz curve and Gini coefficient.

In suicidal studies, inequality indices were used to gauge income inequality finding association between the distribution of suicide and economic status, rather than used as a summary to quantify regional discrepant distribution of suicide (Unnithan and Whitt 1992, Andres 2005). In other words, inequality measures were not used for assessing suicide inequality *per se*, but for assessing inequality of predictor variable, such as economic status. The measurement of health disparity in regional perspective have not been shed light on in studies in public health involving suicide studies.

Instead, geographical studies on suicide have focused on spatial cluster analysis to assess the extent of spatial pattern which comes from uneven geographical distribution. Technically, Moran's I and Geary's c statistics have been mainly used to find global spatial autocorrelation (Middleton et al. 2008), while Kulldorff's scan statistic has been used for exploring the location of cluster (Bando et al. 2012b, Qi et al. 2012).

2.3.2. Spatial Measures for Regional Inequality

Regional inequality has been studied from sociology and (urban) social geography, in the current of measuring residential segregation. Dissimilarity index, proposed by Duncan and Duncan (1955), has served as the most popular segregation index to measure inequality and segregation. After the long period of "*Pax Duncana*", novel ideas for measurement of segregation were proposed (Massey and Denton 1988). Massey and Denton (1988) classified segregation indices into 5 dimensions: evenness, exposure, concentration, centralization, and clustering.[①] Following their classification, dissimilarity index is included in the evenness axis, which measures the uneven

---

[①] Reardon and O'Sullivan (2004) evaluated the evenness and exposure dimensions are aspatial, whilst other three dimensions are plainly spatial in the formation of Massey and Denton (1988).

distribution over areal units, together with popular inequality indices such as Gini coefficient, Theil's entropy index, and Atkinson index.

Dissimilarity index of Duncan and Duncan (1955) began to be criticized for the lack of contemplating spatial arrangement of disparity. White (1983) brought up the *checkerboard problem* to shed light on spatial proximity of neighboring areas in assessing dissimilarity among regions. A wide array of spatial segregation indices started to be developed, incorporating spatial structures of areal units in assessing regional disparity, especially based on spatial arrangement of regions (White 1983, Morrill 1991, Wong 1993, Wong 2002, Wong 2005, Feitosa et al. 2007).

Reardon and O'Sullivan (2004) proposed alternative two dimensions for spatial segregation, aggregating 5 dimensions of segregation suggested by Massey and Denton (1988): spatial evenness (or clustering) and spatial exposures (or isolation) (Figure 2). The spatial evenness integrates the dimensions of evenness and clustering, while spatial exposures integrates the dimensions of isolation and exposure. For measuring spatial evenness, spatial dissimilarity indices have been proposed (Reardon and O'Sullivan 2004, Feitosa et al. 2007).

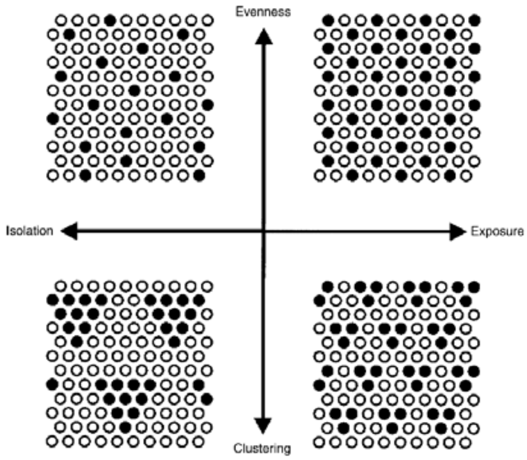


Figure 2. Dimensions of spatial segregation (Reardon and O'Sullivan, 2004: 126)

Ⅲ. Research Methodology

3.1. Categorical Data Analysis

Most of the data used in social science including national mortality data contain categorical variables rather than continuous. Among the wide array of variables, some variables have associations while others not. To reveal complex relationship among categorical variables, methodologies for categorical data analysis have been developed. Some of the techniques were used for enabling to derive thorough description on the categorical data.

3.1.1. Contingency Table Analysis

3.1.1.1. Test of independence

Contingency table is frequently used to explore the categorical data with combinations of variables. Independence test reveals the existence of relationship between those variables. Statistical independence occurs when the population conditional distributions of a variable are identical at each level of the other variable. This is equivalent to the attribute that all joint probabilities are equal to the product of their marginal probabilities (Agresti, 2002: 24).

The null hypothesis of independence can be tested with two goodness−of−fit statistics: *Pearson's chi−squared test statistic* (X^2) and *Likelihood−ratio chi−squared test statistic* (G). Pearsons's test statistic computes the expected counts and compares them to the observed values. With $I \times J$ contingency table, the Pearson goodness—of−fit statistic is defined as following:

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \qquad (1)$$

, where O_{ij} and E_{ij}, respectively, represent observed and expected frequency of ith row and jth column for $i = 1, \dots, I, j = 1, \dots J$.

The log-likelihood test statistic is defined as:

$$G = 2\sum_{i=1}^{I}\sum_{j=1}^{J} O_{ij}\log(\frac{O_{ij}}{E_{ij}}) \quad (2)$$

, where the notations are equivalent to equation (1).

With the degrees of freedom, defined as $(I-1)(J-1)$, the statistic is tested with chi-squared distribution. If null hypothesis rejected with the expected values, two variables are deemed to be independent to each other.

3.1.1.2. Measures of association

When there are large frequencies in each cell in a contingency table, it is natural that null hypothesis are rejected. Rather, it is anomalous that the statistics are not significant. Thus, association measure is needed to investigate how strong the dependencies are. The extent to which the categorical variables are associated is measured with Phi-coefficient (ϕ), contingency coefficient (C), Cramer's V. These association measures are defined respectively as following:

$$\phi = \frac{X^2}{n} \quad (3)$$

$$C = \sqrt{\frac{X^2}{n + X^2}} \quad (4)$$

$$V = \sqrt{\frac{\phi}{\min(I,J)}} \quad (5)$$

, where n is total sum of cell frequencies, and X^2 is Pearson's chi-square statistics. If the contingency table is 2×2 table, Cramer's V is identical to Phi coefficient.

3.1.1.3. Detection of contributing cells to the dependency

Then, where in a table contribute to the dependency between factors? For this analysis, two methods can be used. First, table can be partitioned into several sub-tables and examined respectively, revealing the contributions of each factor on the dependency. While sum of Likelihood ratio statistics of sub-tables is identical to the LR statistics of original table, Pearson Chi-squared statistics is not in that way. That's why LR statistic should be used in partitioning table. It is worth noting the rules for partitioning table. Three principles of partitioning table should be kept, which are the following (Agresti, 2002: 84): The summation of degree of freedom (DF) for the sub-tables must be identical to DF for the full table; Each cell count in the full table must be a cell count in one and only one sub-table; Each marginal total of the full table must be a marginal total for one and only one sub-table.

The second method is to identify standardized Pearson residuals. Chi-statistic is calculated with the summation of squared difference between observed and expected frequency. In examining the difference cell-by-cell, larger expected frequency tend to yield larger difference (Agresti 2007). For comparison, raw residuals should be normalized. Standardized Pearson residuals enable to reveal in which cell the significant difference is shown. For visualization, standardized Pearson residuals are often represented with mosaic plot, which displays marginal distributions of contingency table.

3.1.2. Log-linear Model for Multi-way Contingency Table

Log-linear model, which is of use to model cell counts in contingency table, can be implemented for the analysis of associations and interactions among categories. It can demonstrate a wide array of patterns of independence and association (Agresti 2002). Synthesizing and expanding the tests and statistics of two-way contingency table analysis, log-linear model can be expanded to examining multi-way contingency table. Complex inter-relationship among categorical variables can be modelled and revealed with log-linear model.

Following the notation used in Agresti (2007), with $I \times J$ two-way contingency table, the expected cell frequencies (μ_{ij}) of saturated model can be expressed as:

$$\log(\mu_{ij}) = \lambda + \lambda_i{}^A + \lambda_j{}^B + \lambda_{ij}{}^{AB} \qquad (6)$$

, where λ indicates each coefficient, and categorical variables A and B include $i = 1, \ldots I$ and $j = 1, \ldots J$ levels respectively. If this model is fitted, it implies the existence of association, and the coefficient of the interaction term ($\lambda_{ij}{}^{AB}$) can be interpreted as log-transformed odds ratio.

In examining inter-relationship of categorical variables, model selection is an important, but difficult part. Deviance of the model, which is equal to G-statistic, is commonly used to evaluate the extent to which the model is well-fitted. However, it is not the only criterion for the assessment, because alternative models can be selected with additional criteria of model selection. Partial association test on hierarchical models is one way to select model as described in Legendre and Legendre (2012: 239). This test compares a model which contains interaction term of interest with another model which does not have the partial association term. Even though this is useful in model selection, it is nearly impossible to take account of all hierarchical models because of numerous combinations of categorical variables.② Thus, stepwise selection is suggested by Legendre and Legendre (2012).

In this study, for the first, candidate models which are in simple format will be taken into account for easy interpretation. Then, stepwise selection will be implemented so that the model which has the minimum AIC can be chosen through diagnostic test. The models produced by first two steps will be compared. Selection of single model is a compromise between model performance and simplicity of interpretation. From the more appropriate model, terms will be added or removed manually for simplifying interpretation.

② For three-way contingency table, there are 8 hierarchical models. 113 hierarchical models for four-way contingency table, and so forth. Legendre and Legendre (2012: 239).

3.2. Disease Mapping for Representing Ecological Risk

3.2.1. Age-sex Standardization of Rates

Generally, in regional scale, health risk is expressed by crude rate, which is calculated as the total number of cases divided by the total number of population in a region. It assumes that suicide rates are identical among different demographical strata within a region. In other words, it does not consider population structure at all. In contrast, standardization techniques enhance comparability of region and represent the spatial distribution more accurately, considering the attributes of regions. Standardized rates can be computed through two standardization methods: direct and indirect standardization.

Direct Standardized Rate (DSR) is calculated is calculated with the following formula:

$$DSR = \frac{\sum_{j=1}^{J}(C_j/P_j){P_j}^S}{{P_+}^S} \tag{7}$$

, where C is case, P is population; subscript j represents each category in $1,\ldots J$, while + denotes total value; superscript S means standard.

Standardized Mortality Rate (SMR) is the most common application of indirect standardization. It compares the number of observed cases in the study population to the number of cases expected:

$$SMR = C_+/E_+ \tag{8}$$

, where the expected total number of case (E_+) is derived from the following equation.

$$E_+ = \sum_{j=1}^{J}\frac{{C_j}^S}{{P_j}^S}P_j \tag{9}$$

In fact, indirect standardized rates can be derived by multiplication of SMR and the crude rate in the standard population:

$$\mathrm{ISR} = \mathrm{SMR}\frac{{C_+}^S}{{P_+}^S} \tag{10}$$

Each standardization method has its own purpose and advantages. The primary strength of direct standardization is comparability. In other words, it makes it possible to compare spatial distribution and temporal change of health risk in straightforward manner, because it directly represents the risk of population.

Many international studies have also used the indirect method. The primary reason is that indirect standardization only requires the information on the number of age-specific population in each region, while direct method needs more detailed information on the case, which is often not available. In contrast to direct standardization, SMR cannot assure comparability, only informing excess cases needing health intervention (Pickle 2009). With the expense of the lack of comparability, resulted from the weighting scheme by own strata within region, the indirect method can be used when the total burden of health disease is of interest (Wakefield 2007). Aside from data availability, the strength of indirect standardization is that stratum-specific estimation is stable, while direct standardized rates are vulnerable to the small change of count in the case of rare disease. Moreover, the SMR can be easily interpreted. In addition, its applicability to complex models can be pointed out as another strength. It is usual model that case of regions i follows the Poisson model

$$C_i|\theta_i \sim \mathrm{Po}(E_i\theta_i) \qquad (11)$$

, where θ_i is the true RR of disease in region i. Under Poisson distribution, SMR is the maximum likelihood estimate of true RR in each area (Banerjee, Gelfand and Carlin 2004, Middleton et al. 2008). This implies that it can be extended to more sophisticated models to estimate RR more precisely. To sum up, selecting standardization method depends on the purpose of mapping.

3.2.2. Bayesian Smoothing Technique

In small-scale studies, it is often encountered that counts of disease events are normally low, even with zero counts (Richardson et al. 2004). Especially in health study, low counts of disease is typical because of relatively low incidence rate of disease. Faced with sparse data, standardized mortality rates are often estimated with large standard error. This statistical instability is called "small area problem". With this problem, geographical pattern of RRs can be misrepresented, because inaccurately estimated values dominate spatial pattern (Clayton and Kaldor 1987).

Smoothing techniques have been developed to reduce the instability and provide trustful figures. The basic concept of smoothing is borrowing information from other (or neighboring) regions to estimate more concrete RR (Waller and Gotway 2004). It is a compromise between the rates from each region and global mean, and the extent to which the compromise exist is often referred to as *shrinkage factor*. Anselin, Lozano and Koschinsky (2006) introduced various correction methods to estimate rates more precisely, such as local averaging, spatial filtering, regionalization methods, and Bayesian smoothing. By mapping statistically stable rates, more precise information on the geographical pattern of health risks can be represented. In this study, Bayesian smoothing techniques were implemented to stabilize large variability in estimation. Smoothing techniques derived from Bayesian principles are also known as model-based smoothing (Anselin et al. 2006).

3.2.2.1. Empirical Bayesian approach

Bayesian smoothed risk estimators based on Poisson-gamma model was proposed by Clayton and Kaldor (1987). In Poisson-Gamma model, observed counts of health outcome is expressed as the following two-level model:

$$O_i \sim Poisson(\theta_i E_i)$$
$$\theta_i \sim Gamma(\upsilon, \alpha) \quad (12)$$

With assumption that RRs are independent identically distributed(*iid*), drawn from a gamma distribution with mean ν/α and variance ν/α^2, the posterior expectation of θ_i conditional on observation values is

$$\mathrm{E}(\theta_i|O_i;\ \alpha,\nu) = \frac{C_i + \nu}{E_i + \alpha} \tag{13}$$

, which can be rewritten with clarification of shrinkage factors as following:

$$\mathrm{E}(\theta_i|C_i;\ \alpha,\nu) = \frac{E_i}{E_i + \alpha} SMR_i + (1 - \frac{E_i}{E_i + \alpha}) \frac{\nu}{\alpha} \tag{14}$$

From the first derivative of log-likelihood and asymptotic expectation for the variance of the mixing distribution, following equations are derived:

$$\frac{\hat{\nu}}{\hat{\alpha}} = \frac{1}{\mathrm{N}} \sum_i \frac{O_i + \hat{\nu}}{E_i + \hat{\alpha}} = \frac{1}{N} \sum_i \widehat{\theta_i} \tag{15}$$

$$\frac{\hat{\nu}}{\hat{\alpha}^2} = \frac{1}{\mathrm{N} - 1} \sum_i (1 + \frac{\hat{\alpha}}{E_i})(\widehat{\theta_i} - \hat{\nu}/\hat{\alpha})^2 \tag{16}$$

The parameters α, ν and RR of ith region (θ_i) are estimated with recursive computing with equations (13), (15), and (16).

Another empirical Bayesian smoother was proposed by Clayton and Kaldor (1987) assuming multivariate log-normal distribution for logarithm of the RRs. In this estimation, they did not use simple log RRs, but bias-corrected one:

$$\beta_i = \log(\frac{C_i + 1/2}{E_i}) \tag{17}$$

With EM algorithms updating the estimates of common mean ($\widehat{\Phi}$) and variance ($\hat{\sigma}^2$), empirical Bayes estimate

$$\widehat{\beta_i} = \frac{\widehat{\Phi} + (C_i + \frac{1}{2})\hat{\sigma}^2 \log[\left(C_i + \frac{1}{2}\right)/E_i] - \hat{\sigma}^2/2}{1 + \left(C_i + \frac{1}{2}\right)\hat{\sigma}^2} \tag{18}$$

is also repeatedly updated. The estimated RRs are equivalent to exponential of $\widehat{\beta_i}$.

To incorporate spatial arrangement in the estimation, Marshall (1991) proposed local estimator, which is similar to global estimator proposed by the author, but only the neighboring areas are reflected in calculation. In other words, statistically instable RRs shrink to neighborhood mean, instead of employing global shrinkage. It is based on the postulation that disease rates of closer neighborhood areas are similar and relatively homogeneous (Marshall, 1991).

3.2.2.2. Hierarchical Bayesian smoothing

In Bayesian hierarchical models, little prior information is introduced in estimation. Technically, υ and α in formula (13) are estimated from vague gamma priors in full Bayesian estimation. Two hierarchical models were applied in this study. First, Bayesian hierarchical Poisson−gamma model was used to estimate RRs of regions. It can be recognized as global smoothing technique due to the lack of consideration of spatial structure. In opposite, the second method takes spatial dependency into account, formulating second stage model as following (Best, Richardson and Thomson 2005):

$$\begin{gathered} O_i \sim Poisson(\theta_i E_i) \\ log\theta_i = V_i + U_i \\ V_i \sim N(0, {\sigma_v}^2) \\ U_i | U_{(-i)} \sim N(\frac{\sum_j w_{ij} u_j}{w_{i+}}, \frac{{\sigma_u}^2}{w_{i+}}) \end{gathered} \qquad (19)$$

, where spatial random variable U_i follow conditional autoregressive model, w_{ij} denotes a spatial weight between regions i and j, and ${\sigma_v}^2$, ${\sigma_u}^2$ indicate variance. This model was proposed first by Besag, York and Mollié (1991), thus being called as BYM model very often.[③] In this model, the variability of logarithmic RRs is divided into a spatially correlated variable and an area−independent effect. This conditional autoregressive model permits to include covariates to explain part of variability of the RRs. Best et al. (2005) found the usefulness of BYM model with and/or without covariates through simulation test.

③ The model is also often called as CAR model due to the specification of conditional autoregressive model for spatial random variable.

3.2.3. Cartogram

Cartogram can be used as an alternative method of representing spatial information. In contrast to conventional mapping which preserves area (planimetric map), cartogram transforms area, shape, and topology to represent statistical information. Hennig (2011) categorized three common types of cartograms: Dorling and Dorling-like cartograms, non-contiguous cartograms, and contiguous cartograms.④ He pointed out that contiguous cartograms can be recognized as a type of map projections and used as a basemap. He suggested to produce cartograms based on population.

Following the suggestion, contiguous cartograms were used as a map projection in this study to represent additional information. Cartograms were produced, however, based not only on the population, but the values of absolute number of cases. Andresen et al. (2009) suggested the use of cartogram in crime mapping. In the study, cartograms were used in mapping rate and location quotient of violent crime in this study, to give additional information on the absolute count and the risk of violent crime, respectively. In this case, the observed case, numerator, was used to scale the areas of each region. Cromley and Cromley (2009) otherwise suggested to transform the spatial unit to be proportional to the denominator, population, to improve the communication of the health risk. In this study, cartograms were made based on the values of both numerator and denominator.

For producing contiguous cartograms, diffusion-based method, developed by Gastner and Newman (2004), was used in the transformation (Figure 3). ArcGIS 10.1 software and the Cartogram Geoprocessing Tool, an ArcScript tool established by Tom Gross in 2009, were used for implementing.

---

④ Geographical units are represented differently by each type. It is represented by circles in Dorling cartogram. For the others, the size of units are rescaled. While non-contiguous cartograms preserve shape of each unit, contiguous cartograms do not. However, contiguous cartograms preserve the contiguity of geographical units.

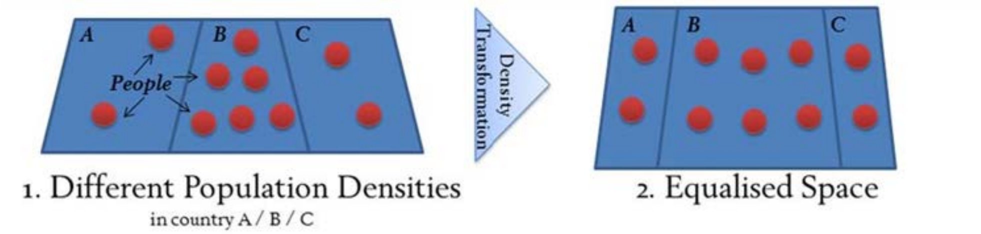


Figure 3. Diffusion process in a cartogram transformation (Hennig, 2011: 26)

3.3. Measurements of Regional Inequality

3.3.1. A-spatial Measurement of Regional Disparity

3.3.1.1. Extended Gini index

Gini index is the most widely used concepts in measuring inequality. Usually, Gini index is defined as the normalized area between the Lorenz curve and the diagonal line in the plot of Lorenz curve. The curve represent the observed distribution of health outcomes, while the diagonal line means theoretical state of complete evenness. The generalization of Gini coefficient was proposed by Yitzhaki (1983), which is also called as extended Gini index or S-Gini. This index extended Gini index in covariance form:

$$G(\nu) = -\nu\, cov(\frac{x}{\mu}, (1 - F(x))^{\nu-1} \qquad (20)$$

, where ν is inequality aversion parameter. The extended Gini index ($G(\nu)$) is defined for $\nu > 1$, and the coefficient with larger ν is more sensitive to the change of lower tail in the distribution of x. When ν=2, $G(\nu)$ is equal to the original Gini index.

3.3.1.2. Generalized Entropy Index (GEI)

Generalized Entropy Index, which is based on information theory, is defined as following:

$$GE(\theta) = \frac{1}{\theta^2 - \theta}[\frac{1}{n}\sum_{i=1}^{n}(\frac{x_i}{\mu})^{\theta} - 1] \quad (21)$$

, where the parameter θ denotes the weight given to distances between income (Litchfie, 1999: 3), and μ denotes the average of all x values in n regions. Lower the value of parameter θ, greater weight is given to lower tail of distribution. The frequently used values of parameter θ are 0, 1, and 2. GE(0) is mean logarithmic deviation, GE(1) is Theil's index, GE(2) is square root of coefficient of variation.

Generalized Entropy Index meets the set of axioms suggested by Cowell (2000),⑤ including decomposability. It is superiority of GEI, because only this index can be completely decomposed into between-group ($I_{between}$) and within-group inequality (I_{within}) in additive form, which means that the summation of the two components is always equal to the total inequality.

3.3.1.3. Dissimilarity Index

Inequality can also be measured with Duncan and Duncan (1955)'s dissimilarity index (D), which is defined as following:

$$D = \frac{1}{2}\sum_{i=1}^{n}|\frac{X_i}{X} - \frac{Y_i}{Y}] \quad (22)$$

, where X and Y denotes two different compositional group for region i ($i = 1, \dots I$). Mathematically, it is equivalent to the maximum distance between the Lorenz curve and diagonal line. This index is often be interpreted as the proportion of minority group that would need to be transferred from regions to achieve complete evenness.

⑤ The axioms, required for inequality measures, were suggested as following: Pigou-Dalton transfer principle, income scale independence, principle of population, anonymity, decomposability (Litchfield, J. A. (1999) Inequality: methods and tools. *Text for the World Bank PovertyNet.*)

3.3.2. Spatial Cluster Analysis

The purpose of spatial cluster analysis is to reveal whether spatial pattern of study objects exists, and if it does, to identify the location where the pattern is present. It is an essential step for exploratory analysis, because more in-depth investigation can be brought about only after spatial cluster pattern is identified (Wang 2004).

3.3.2.1. Global Detection of Cluster

To evaluate the existence and the extent to which spatial autocorrelation, global indices have been used. Moran's I is one of the most popular global index (Moran 1948):

$$I = \frac{n \times \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (Y_i - \bar{Y})^2} \quad (23)$$

As shown above, Moran's I measures the similarity of values which are deviated from the mean. However, Moran's I power remained same, irrespective of heterogeneous population spatial structure (Assuncao and Reis 1999). And the equation (23) is not proper to analyze spatial correlation of count data with populations. Moran's I have been modified in several ways to address this problem. Walter (1992) modified the equation as following:

$$I_{cr} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{C_i - rP_i}{\sqrt{rP_i}} \frac{C_j - rP_j}{\sqrt{rP_j}}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \quad (24)$$

, where C is the number of case, P is population, r is constant risk, defined as $\frac{P}{C}$, and w_{ij} is a spatial weight matrix between ith region to jth regions.

Tango (1995) proposed a test, dubbed as Tango's excess events test (EET), generalizing his prior suggestion on the detection of temporal clustering (Tango 1984).

$$T = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(\lambda) \frac{C_i - rP_i}{C} \frac{C_j - rP_j}{C} \quad (25)$$

, where $w_{ij}(\lambda) = \exp((-4(\frac{d_{ij}}{\lambda})^2)$ and expected value of region i and j (e_i, e_j) is defined. λ can be recognized as *cluster size* and equal to the maximum distance between observations (Tango, 2000: 193).

Rogerson (1999) suggested to use spatial chi-square statistics of equation (27) with the weight matrix

$$w_{ij} = \frac{a_{ij}}{\sqrt{\frac{P_i}{P} * \frac{P_j}{P}}} \qquad (26)$$

The equation (27) can be rewritten as following:

$$\begin{aligned} R &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(\frac{C_i}{C} - \frac{P_i}{P})(\frac{C_j}{C} - \frac{P_j}{P}) \\ &= \sum_{i=1}^{n} \frac{(\frac{C_i}{C} - \frac{P_i}{P})^2}{\frac{P_i}{P}} + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \frac{(\frac{C_i}{C} - \frac{P_i}{P})(\frac{C_j}{C} - \frac{P_j}{P})}{\sqrt{\frac{P_i}{P} * \frac{P_j}{P}}}) \end{aligned} \qquad (27)$$

In the equation, the first term corresponds to the aspatial chi-square statistic, while the second term measures the spatial dissimilarity between regions.

In this study, the expected number of cases in region i and j (rP_i, rP_j) from equations (24) and (25), were replaced by the expected number of counts, estimated by standardization to reflect different sub-group population structures of each region. This expected values stand for more concrete constant risk, taking into account age-sex strata.

3.3.2.2. Local Detection of Cluster

For detecting individual clusters, local measures of similarity were developed which were termed as local indicators of spatial autocorrelation by Anselin (1995).

The most popular local Moran' I is defined as following:

$$I_i = (Y_i - \bar{Y})\sum_{j=1}^{n} w_{ij}(Y_j - \bar{Y})) \qquad (28)$$

Under the constant risk hypothesis, local version of Moran's I statistic can be adjusted considering heterogeneous population structure following (Waller and Gotway, 2006: 239):

$$I_{i,cr} = \frac{(C_i - rP_i)}{C_i} \sum_{j=1}^{n} w_{ij} \frac{(C_j - rP_j)}{C_j} \quad (29)$$

Furthermore, Local version of Tango's EET was proposed by Rogerson (1999). For a given point i, the test statistic is

$$R_i = \frac{1}{\sqrt{P_i/P}} \left(\frac{C_i}{C} - \frac{P_i}{P}\right) \sum_{j=1}^{n} w_{ij} \left(\frac{\frac{C_i}{C} - \frac{P_j}{P}}{\sqrt{P_j/P}}\right) \quad (30)$$

, where w_{ij} is adjacency matrix which was identical to equation (26).

Spatial scan statistics introduced by Kulldorff and Nagarwalla (1995) and Kulldorff (1997) is another method to reveal local cluster. A number of overlapping circles, which are possible contender for being a cluster, are used in this scanning method. The alternative hypothesis is that risk within a circular window Z is higher compared with outside it. Under the Poisson distribution, the test statistic of the likelihood ratio maximized for each Z is defined as following (Tango, 2010: 80):

$$\lambda = \sup_{z \in Z} \left(\frac{C(Z)}{e(Z)}\right)^{C(Z)} \left(\frac{C - C(Z)}{C - e(Z)}\right)^{C - C(Z)} I\left(\frac{C(Z)}{e(Z)} > \frac{C - C(Z)}{C - e(Z)}\right) \quad (31)$$

, where C(Z) is the number of cases within Z, and I is the indicator function. The most likely cluster (MLC) is defined as the window which records maximum likelihood (Kulldorff 1997).

3.3.3. Spatial Dissimilarity Indices

These spatial measurements of inequality, developed in the field of studying residential segregation, can be applied in epidemiological geography to assess regional inequality of suicide, with the assumption of increasing relative region inequality which comes from clustering pattern of suicide.

In the current of spatial turn in residential segregation studies, Morrill (1991) adjusted D index (Duncan and Duncan, 1955) to distinguish various patterns of distribution over space. The adjacent relationships of regions were incorporated as spatial interactions in the index. For this reason, this index is called the boundary modified D index (D_{adj}). The index is defined as following:

$$D_{adj} = D - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|w_{ij}(X_i - X_j)|}{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}} \tag{32}$$

In the calculation of spatial dissimilarity index of Reardon and O'Sullivan (2004), spatial structures and interactions among data points are reflected by population density of the local environment, which express characteristics of a locality. The use of population density function was emphasized to resolve Modifiable Areal Unit Problem (MAUP). The spatial dissimilarity index ($\widetilde{D}$) is defined as following (Reardon and O'Sullivan, 2004: 140):

$$\widetilde{D} = \sum_{m=1}^{M}\int_{p\in R}\frac{\tau_m}{2NI}|\tilde{\pi}_{pm} - \pi_m|dp \tag{33}$$

, where

$$I = \sum_{m=1}^{M}(\pi_m)(1 - \pi_m) \tag{34}$$

$$\tilde{\pi}_{pm} = \frac{\Phi_p \int_{q\in R}\tau_{qm}\phi(p,q)dq}{\Phi_p \int_{q\in R}\tau_q\phi(p,q)dq} \tag{35}$$

In the equations above, π_m denotes the proportion of group m of total population, $\tilde{\pi}_{pm}$ is defined as the proportion in group m in the local environment of point p, of which total number is N. Lastly, $\phi(p,q)$ represents spatial proximity between the point p and q.

The formula is based on complete individual point data. Because individual point data are not available in most cases, they suggested spatial smoothing of population as an alternative approach. In this study, kernel density estimation was used for spatial smoothing.

Feitosa et al. (2007) adopted the ideas of Reardon and O'Sullivan (2004) to areal unit data, with replacing population density to local population intensity, which is a weighted average of population counts in a locality. As mentioned in their article, the local population intensity is sub-concept to the composite population counts which was proposed by Wong (2005), which is defined as:

$$cp_{mi} = \sum_{j\,(=i)} w_{ij}(p_{mj}) \quad (36)$$

Spatial dissimilarity were measured among each areal unit, not among individual point data or estimated population density surface. The formula of spatial dissimilarity index ($\breve{D}(m)$) is following (Feitosa, 2007: 304−305):

$$\breve{D}(m) = \sum_{j=1}^{J} \sum_{m=1}^{M} \frac{N_j}{2NI} |\breve{\tau}_{jm} - \tau_m| \quad (37)$$

, where

$$I = \sum_{m=1}^{M} (\tau_m)(1 - \tau_m) \quad (38)$$

$$\breve{\tau}_{jm} = \frac{\sum_{j=1}^{J} k(N_{jm})}{\sum_{j=1}^{J} k(N_j)} \quad (39)$$

In equation (36), τ_m is the proportion of group m in the study area; $\breve{\tau}_{jm}$ is the local proportion of group m in locality j; $k()$ represents kernel estimation.

Feitosa et al. (2007) also introduced local version of spatial segregation index which decomposes the global indices to identify the extent to which each locality contributes to the global spatial segregation. The local spatial dissimilarity index was formulated as following:

$$\breve{d}_j(m) = \sum_{m=1}^{M} \frac{N_j}{2NI} |\breve{\tau}_{jm} - \tau_m| \quad (40)$$

, where notations are identical to those of equation (33) to (35), except that the values are calculated for each region j.

Ⅳ. Analysis and Results

4.1. Explorative Analysis on Suicide Mortality Data

4.1.1. Data Description

It is typical in epidemiological geography to examine the number of deaths, taking into account of the total population and the composition, as an assessment of public health (Shaw, Dorling and Mitchell 2002). This study focus on complete suicide and utilize national mortality data to examine geographical distribution and pattern.⑥

The national mortality data for the five years of study period (2008−2012) were obtained from the Korean National Statistics Office (KNSO) with Micro Data Service System (MDSS). The data is established based on the death certificate, containing various kinds of information, such as address, sex, educational level of the dead, time and place of occurrence, and the cause of death.

From 1,281,408 cases of total mortality, 74,084 suicidal cases were extracted for the study period. And 181 cases which were conducted by foreigners were excluded. Finally, 73,903 cases were selected to be analyzed in this study. Only "intentional self−harm", coded as X640−X800 in the sixth revision of Korean Standard Classification of Diseases (KSCD−6)⑦, is referred to suicide and analyzed in this study.

It is worth noting that the mortality data was analyzed in Si−Gun−Gu (SGG) level⑧ for spatial analysis. This geographical scale

---

⑥ Suicide behavior is divided into suicide ideation (suicidal thought), suicide attempts (parasuicide), and complete suicide. Analyses on suicide ideation and suicide attempts are implemented with survey data available which include data in the survey format. In the contrary, studies on complete suicide are primarily depend on incidence or mortality data.

⑦ KCD-6 follows the standard and system of Tenth International Classification of Diseases (ICD-10)

⑧ Administrative districts in Korea are defined in three levels: Si-Do, Si-was used in this analysis because administrative codes of the addresses of complete suicides in the data were only provided in SGG units. Given that mortality data are treated as personal information and protected by law in Korea, the release for finer level is prohibited. SGG is the finest unit available now; thus, this geographic scale was used throughout this study.

Adjusting administrative code is also worth to be mentioned. Administrative code of address were adjusted reflecting temporal changes of the data. In studying with time-series data, it is important to match classifications of categories through the study period to compare them longitudinally. Three administrative units were combined in 2010: Masan-si and Jinhae-si were consolidated into Changwon-si. Merged into one city, the three districts were rather divided and provided more specific information: Changwon-si was divided into Uichang-gu and Seongsan-gu; MasanHoewon-gu and Masanhappo-gu were divided from Masan-si; Jinhae-si became Jinhae-gu. Thus, these districts after 2010 were recoded to be compared with the previous administrative units. Moreover, two administrative units in 2012 were not matched to the previous ones, so modified: Sejong-si and Dangjin-si were linked to Yeongi-gun and Dangjin-gun, respectively.

The cause of death statistic consists of many categories. Table 2 shows the categories employed in this study. The demography category provides individual's information which enables to typify suicides. These categories cannot be arbitrarily chosen or easily modifiable after a suicide has intention of self-harm. Otherwise, method of suicide and time of occurrence are highly depend on the choice of the one who commit suicide.

Some categories were defined or reclassified based on previous studies. In many suicide studies, people aged above 9 is regarded as population at risk, and is classified into 4 categories: young people, aged 15-24; Young adults, aged 25-44; middle-aged, aged 45-65;

Gun-Gu, and Eup-Myeon-Dong. As the medium level, the number of Si-Gun-Gus was 249 in 2008, with 393㎢ median size of area.

elderly people, aged above 65 (Saunderson and Langford 1996, Pearce, Barnett and Jones 2007, Qi et al. 2010, Kim et al. 2011a). Suicide method was classified into 6 categories according to KSCD-6: drowning (X71), fire (X75-X77), hanging (X70), jumping from a high place (X80), poisoning (X60-X69), and others (X72-X74, X78-X79, X81-X84). Poisoning was further divided by sources: gas (X66-X67), pesticide (X68), medication (X60-X64), and other agents (X65, X69). Moreover, suicide means were classified into violent suicide methods and non-violent ones. Conventionally, non-violent suicidal methods have been defined as poisoning (Chen et al. 2009, Ajdacic-Gross et al. 2010, Ahn et al. 2012). For holiday, two big holidays were selected: New Year's Day in lunar calendar and Chuseok. Lastly, a female actress, who killed herself in October 2, 2008, was selected to explain the effect on the suicide engendered from the celebrity's suicide.

Table 2. Categorization of suicides on socio-demographic characteristics, suicide method, and time of occurrence

	Category	Sub-categories
Socio-Demography	Sex	Male / Female
	Age	Child(<15) / Young(15-24) / Young Adult (25-44) / Middle Aged (45-64) / Elderly (65+)
	Marital Status	Unmarried / Married / Divorced / Bereaved
	Job	Manager / Professional / Officer / Service / Agriculture and Fishery / Technical / Machinery / Unskilled / Unemployed
	Address (Si-Do)	Seoul / Busan / Daegu / Incheon / Gwangju / Daejeon / Ulsan / Gyunggi / Gangwon / Chungbuk / Chungnam / Jeonbuk / Jeonnam / Gyungbuk / Gyungnam / Jeju
Method	All method	Drowning / Fire / Hang / Jump / Poisoning (Gas, Pesticide, Medication, Other agent) / Others
	Violent method	Violent : Drowning / Fire / Hang / Jump / Others Non-violent: Poisoning
Time	Season	Spinrg / Summar / Fall / Winter
	Weekday	Sunday / Monday / Tuesday / Wednesday / Thursday / Friday / Saturday
	Holiday	3-days before major holiday / Major holiday / 3-days after major holiday
	Celebrity	28-days before celebrity's suicide / 28-days after celebrity's suicide (including the day)

For population data, age- and sex-specific population data in the corresponding geographic level of the suicide data[9] were also derived from the KNSO for the identical period.

4.1.2. Descriptive Statistics

Table 3 represents the frequencies of suicide classified by socio-demographic characteristics and suicide method within each year of study period. Note that total number of suicide in each categories can be different with each other due to the lack of values. With the total number of suicides in each year, it can be acknowledged that the number of suicides decreased in 2012, while it had increased consistently until 2011. It is also shown that suicide of male outnumbers that of female. In educational level and marital status category, the proportions are the highest in young adults and secondary educational level. It is notable that unemployed people take more than a half of total suicide. This should not be taken at face value, however, because soldiers, students, and housewives come under the unemployment category in the data from KNSO. It means that the actual number of suicides committed by the people who lost their jobs might be much less.

Table 3. Frequency table of suicide in Korea

Category	Sub-category	2008	2009	2010	2011	2012	Sum
Sex	Male	8260	9936	10329	10866	9622	49013
	Female	4598	5477	5237	5040	4538	24890
	Sum	12858	15413	15566	15906	14160	73903
Age	Young	318	450	353	373	337	1831
	Young Adult	4876	5790	5534	5622	4864	26686
	Middle Aged	4103	5092	5293	5505	4935	24928
	Elderly	3561	4071	4378	4406	4023	20439
	Sum	12858	15403	15558	15906	14159	73884

[9] Median population of SGG in 2011 was 145,000 persons.

Marital Status	unmarried	3556	4173	4070	4273	3808	19880
	married	6078	7222	7453	7471	6609	34833
	divorced	1496	1976	1985	2166	1923	9546
	bereaved	1722	2011	2009	1955	1783	9480
	Sum	12852	15382	15517	15865	14123	73739
Job	manager	151	237	284	259	340	1271
	professional	382	625	603	560	568	2738
	officer	733	979	997	1094	801	4604
	service	1090	1372	1365	1333	1358	6518
	Agriculture /fishery	1090	1163	1184	1199	991	5627
	technical	349	384	341	385	379	1838
	machinery	148	201	201	208	247	1005
	unskilled	422	552	525	564	670	2733
	unemployed	7896	9368	9571	9706	8404	44945
	Sum	12261	14881	15071	15308	13758	71279
Cause	Drowning	460	500	429	413	467	2269
	Fire	88	91	73	116	71	439
	Hang	6441	8259	8549	8347	7150	38746
	Jump	1795	2123	2159	2328	2345	10750
	Others	432	394	350	313	268	1757
	Poison	3642	4046	4006	4389	3859	19942
	Sum	12858	15413	15566	15906	14160	73903

Figure 4 illustrates age– and sex–specific rates of suicide during the study period. Discernable difference of suicide rates between genders is not found in the class of aged under 20. Except this age group, however, suicide rates of male considerably outstrip those of female. Moreover, suicide rates display a tendency to increase as the level of age group is getting higher. Even though exceptional pattern is shown in female young adults, having recorded higher suicide rates than middle–aged females. The suicide rates of both age groups have been almost same in recent times. It is worth noting that suicide rates of middle–aged males are even greater than the rates of elderly women. This implies that sex is more useful factor when disentangle suicides than age group, while it is also an important factor to be considered. Furthermore, it should be mentioned that the fact that elderly men is the most vulnerable group to suicide is revealed. The suicide rates in the age group shows remarkably high, which are more than three times, compared to the average suicide rate during the study period, 27 suicides per 100,000 people. Chen et al. (2012b) pointed out that unprepared aging and rapid collapse of the traditional values and family system can be pointed out as the putative reasons for the surge of suicide rates found in elderly people in Korea.

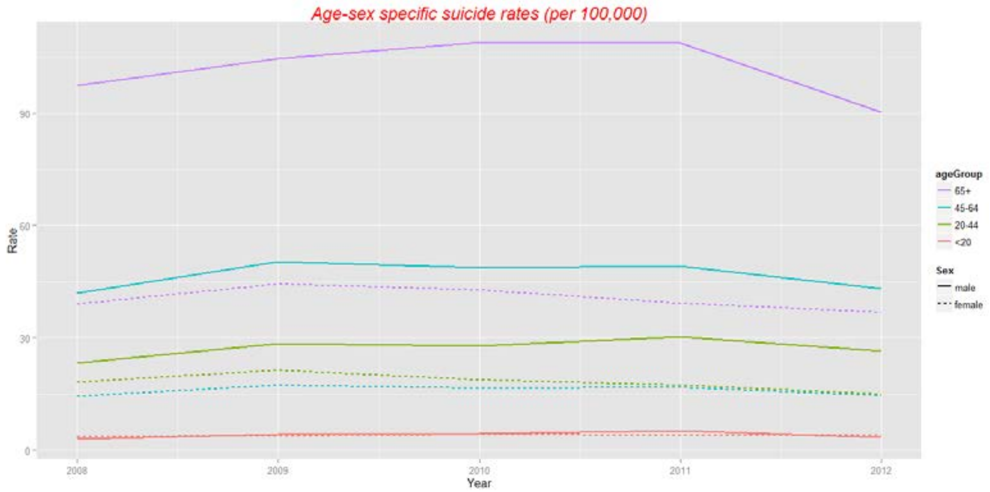


Figure 4. Age-sex specific suicide rates during the study period (per 100,000)

Figure 5 shows the seasonal pattern of suicide in consecutive years. Totally, suicide occurs in spring most frequently, but the difference is not substantial compared to summer. Especially in 2010 and 2012, suicide occurred more in summer than in spring. It is remarkable feature that exceptional pattern is shown in 2008, with largest amount of suicide counts in fall season.

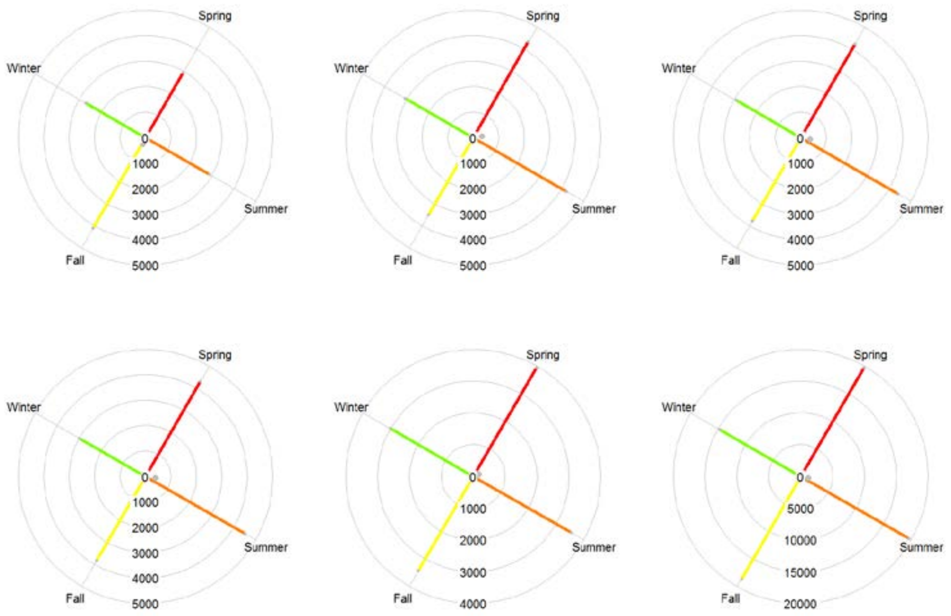


Figure 5. Radial plot of seasonal suicide in 2008 (top-left); 2009 (top); 2010 (top-right); 2011 (bottom-left), 2012 (bottom), and for five years (bottom-right)

To identify temporal pattern more in detail, the frequencies of monthly suicides are represented in Figure 6. The monthly aggregated cases are illustrated in rose plot, which is useful to the circular characteristics of temporal data proficiently, and histogram. In the upper left panel, October seems to be the most perilous month in the study period. However, it is revealed that May and June is much more hazardous than October when year 2008 is removed. The result is shown at the right panel, showing that October 2008 was significant aberration. It matches to the exceptional pattern found in seasonal variation in Figure 5. The general seasonal pattern is shown that the occurrence of suicides increases until late spring and diminishes gradually. Figure 6 also shows the monthly frequency of suicide, but monthly suicide of each bar is decomposed with sex and age by each year. Similar patterns are appeared in both sex, which alludes insignificant interaction between gender and temporal pattern. An exceptional pattern of year 2008 was common in both male and female. Particularly, substantial increases are shown in October, despite the lowest level of frequencies in year 2008. The number of suicides surged among young adults, and increase rates of female seems to be higher than that of male. Other age groups also presented considerable increase during the same period.

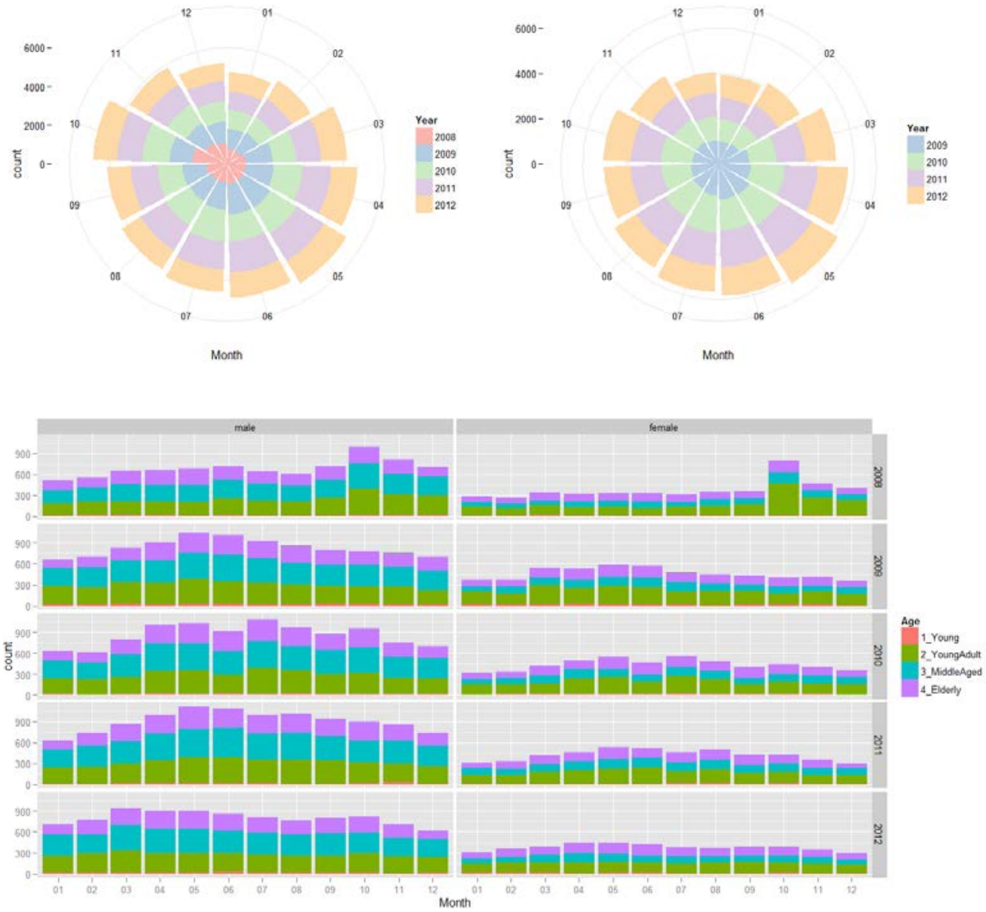


Figure 6. Plots for representing monthly suicide: Rose plot, colored by frequencies of each year (top-left); Rose plot, except year 2008 (top-right); Histogram, colored by age class, on the dimension of year and sex (Bottom)

Figure 7 demonstrates frequencies of suicide in both weekly and daily intervals.[10] Generally, suicide decrease during the weekend and substantially increase and peaks on Monday (Durkheim 1952, Gabennesch 1988, Ajdacic-Gross et al. 2010). As shown in left panel,

[10] No suicide was recorded in 427 days (23.4%) for the study period.

weekly pattern of suicide in Korea conforms to previous studies. In the right panel, rough seasonal pattern is also shown as presented previously. Likewise, aberrant pattern is shown as well. Sudden increase is observed in October 2nd, recording more than double, compared to the previous day. Moreover, the incidence of suicide reached the acme on the 7th day of October, recording 86 cases. In fact, this is the day when suicide occurred the most frequently in the whole study period. Considering the fact that overall frequency level of suicide in 2008 was the least during the study period, this is a remarkable figure.

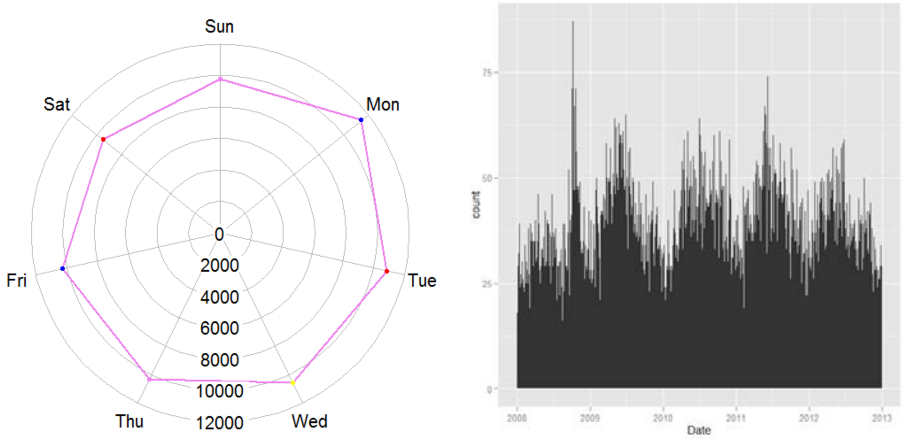


Figure 7. Radial plot of weekly suicide (left); Daily frequencies of suicide (right) for the study period

4.1.3. Contingency Table Analysis on Suicide Data

By explaining the relationships between variables, more interesting facts can be drawn from the data. For this purpose, contingency table analysis was implemented to identify association among variables. All of the combinations cannot be treated in this analysis due to the limit of time and space. Thus, only several relationships among variables will be examined based on previous studies for deriving some meaningful implications.

4.1.3.1. Violent Suicide Means

Many studies on suicide focused on the difference of suicide method selection between male and female. One of the findings is that females are more likely to select non-violent suicide methods when they commit suicide (Denning et al. 2000). Due to the greater success rate, the choice of lethal methods is stated as a plausible reason for the higher rate of completed suicide in males, compared to that of females.

As Callanan and Davis (2012) stated in their review paper, disputes exist on this issue. A previous study, which intensively reviewed articles on suicide method, argued that female tend to use lethal methods in Asia (Wu, Chen and Yip 2012). According to the authors, only in western countries, the previous knowledge on the selection of violent suicide means is relevant. Moreover, a recent comparative study asserted that the dichotomy, Western and non-Western, is not an appropriate approach. In this study, it was found that almost half of female used violent suicide methods in Taiwan and U.S., whereas only less than a third of female used lethal means in Korea and Sweden (Chen et al. 2009).

Wu et al. (2012) stated that poisoning was the leading method in Korea before 2003, but after then, the rankings of common method changed in 2006 and have not changed from then: the order from the first is hanging, poisoning, jumping from the height. Figure 8 demonstrates bar plots representing the proportions of suicide methods by each year, conditioned on sex. As shown in the plot, the pattern did not change yet. Furthermore, contrary to Western case, the proportion of selecting non-lethal method is higher among males than females. Particularly, the proportion of gas poisoning is getting larger from 2008, and this pattern is also shown among females. It is worth noting that suicide cases, committed by jumping from high place, had been considerably increased and reached to the level of poisoning in 2012.

Figure 8. Proportions of suicidal methods for each year, conditioned on sex

To identify the dependency between sex and selection of lethal suicide methods and the extent to which both variables are related, contingency table analysis was utilized. Table 4 is a two−way contingency table of sex and choice of lethal suicide means.

Table 4. Contingency table of sex and violent suicidal method

	Violent suicide means	
Sex	Non−violent	Violent
Male	13,923	35,090
Female	6019	18,871

Pearson's chi−squared statistic is 149.51 (df=1, p<0.0001), and likelihood ratio statistic is 151.28 (df=1, p<0.0001). Thus, it can be said that two variables are not independent. Weak association is resulted, however, between sex and lethal suicide method through the association measures: phi−coefficient, contingency coefficient, and Cramer's V were calculated, and the values were all 0.045.

Furthermore, it should be noted that females were more inclined to select violent suicide methods than males during the study period. Table 5 shows standardized residuals, derived from Pearson's chi−squared test. And Figure 9 is a mosaic plot which visualizes this information graphically. With these results, it can be concluded that the odds of selecting non−violent methods by female is larger than that of male in case of Korea.

Table 5. Standardized residuals of Pearson's Chi−squared test

	Violent Suicide means	
Sex	Non−violent	Violent
Male	12.23	−12.23
Female	−12.23	12.23

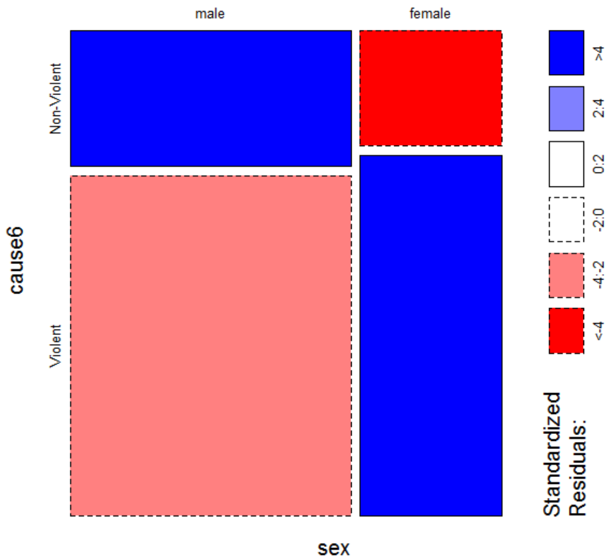


Figure 9. Mosaic plot of sex and violent suicide means with standardized Pearson residuals

Log−linear model for the independence of two variables was fitted, and the model produced identical results conforming to the contingency table analysis. Given Table 4, the result is shown in the Table 6 (Model 1). Note that model expression is followed by Agresti (2002).[11] The deviance of the Model 1, independence model, is 4495, which is equal to the log likelihood ratio statistic of the contingency

[11] Independent relationship is expressed as separate parentheses, whereas variables in the parenthesis represent the existence of dependency between them. For example, (A)(B) means that variable A and variable B is not do not have any association.

table. With 1 degree of freedom, the fitted model cannot reject the null hypothesis of independency, implying the dependency between sex and violent suicide method exists. Model 2 is saturated model for two variables, including the interaction term between two variables. The coefficient of the interaction term is 0.21833, which can be interpreted as log odds ratio of two variables.

Table 6. Log-linear models for sex and violent suicide method

		Model 1	Model 2
Call		(S) (V)	(S, V)
Coefficient	(Intercept)	9.48992	9.54130
	S[F]	−0.67762	−0.83862
	V[V]	0.99543	0.92437
	S[F]:V[V]		**0.21833**
Deviance		151.28	9.91e−13
Df		1	0
Pr(>Chisq)		0	1
AIC		203.2	53.91

Note: S – Sex / V – Violent suicide method
S[F] – Female level of sex category
V[V] – Violent level of violent suicide method category
Bold: logarithmic odds ratio between two variables

Table 7. Three-way contingency table of sex, age, and violent method

		Violent suicide method	
Sex	Age	Non−violent	Violent
Male	Young	75	957
	Young Adult	3,756	12,596
	Middle aged	4,902	13,749
	Elderly	5,189	7,771
Female	Young	59	740
	Young Adult	1383	8,951
	Middle aged	1571	4,706
	Elderly	3006	4,473

Additional dimension can be considered to analyze the dependency between the variables more accurately. Technically, by stratification of contingency table, conditional independence can be tested. As represented in Table 7, additional dimension of age group was included for the stratification.

When Mantel−Haenszel chi−squared test is implemented, the null hypothesis, common odds ratio is equal to 1, is rejected (X^2=141.5, df=1, p−value<0.001). The estimated common odds ratio is 1.244 (CI: 1.20−1.29). However, odds ratios can be heterogeneous by each age strata. With the result from Woolf's test, the assumption of homogeneity is violated (X^2=243, df=3, p−value<0.001). It means that odds ratios differ by each age strata. The calculated odds ratios which are log−transformed are shown in the left panel of Figure 10, whereas the association plot conditioned on each age stratum is represented in the right panel. From the figure, significant association between gender and the choice of violent suicide method is found only in young adult. In this age group, the odds ratio of female selecting lethal suicide methods is 1.96. In the other age bands, sex did not act as an influential factor for selecting violent method.

It is noteworthy that these relationship can also be formulated and explained by log−linear model. For example, Mantel−haenszel test can be formulated as $(S,V)(V,A)$, where S, V, and A respectively denote sex, violent methods, and age variables. As shown in these examples, log−linear model is a generalized method to examine the relationship between categorical variables.

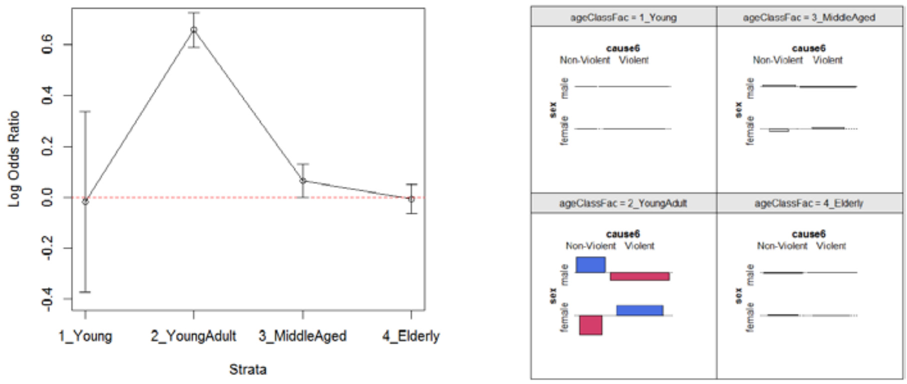


Figure 10. Plot of log odds ratio by age class strata with confidence intervals (left); Association plot of sex and violent suicide method conditioned to age class strata (right)

4.1.3.2. Holiday effect

Holiday has been pointed out as a factor affecting temporal variation of suicide. Before and after holiday, similar patterns are shown in previous studies (Jessen and Jensen 1999, Jessen et al. 1999, Ajdacic−Gross et al. 2010, Nishi et al. 2000). The pattern is that the occurrences of suicide diminish before holiday, and rise again after holiday.

To explain the temporal variation of suicide, the concept of 'broken−promise effect' was suggested (Gabennesch 1988). The effect means that affective events including holiday not only enables to deviate from negative contemplation of potential suicide by raising their expectation, but widening the between raised anticipation and harsh, tenacious reality. Jessen et al. (1999) examined the effect of broken−promise effect and argued that postponing effect engendered by holiday was disclosed. Especially, large fluctuation was discovered in major holiday such as Christmas and Easter, compared to minor holiday.

In another context, a study in Korea defined holiday as a contributor to the social integration, as stated by Durkheim (1952), increasing the interaction between family members (Han, 2008). In this study, decreasing number of suicide was expected before and during the holiday. While significant reduction in elderly people was found, difference from marital status was not shown in the study.

Figure 11 shows daily frequencies of suicide for three days before and after the major and minor holidays. The temporal pattern discovered in previous studies is also shown here, and the variation seems to be larger in major holidays than in other holidays.

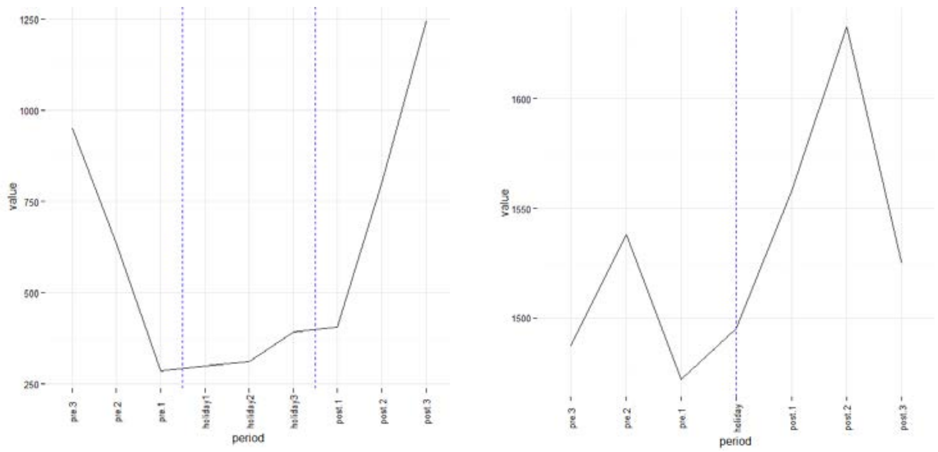


Figure 11. Daily suicide during 3 days before and after the major holidays (left); minor holidays (right)

For developing log-linear model, sex, age, marital status was included in the model as variables. In this analysis, the age of suicide was limited to the people aged over 20 to remove confounding levels, to be free from the structural zero problem in the people aged under 20. If these zero counts are included, it might undermine the relationship inferred from the output of model. Reference factor levels of each variable used in developing models were young adult, male, married, and the days before holiday.

Table 8 shows the fitted log-linear models, and the final model is selected to explain the association among variables. Model 1 is mutual independence model, which has the most basic form. Because this model does not fit well, homogeneous association model was considered (Model 2). A three-way interaction term, (S, A, M), was included in the model because these three variables can be considered as predictor variables in the model. This model fits the best, in terms of AIC. If the models are fitted with stepwise function, Model 2 is selected as the best model. However, it is needed to modify the result of stepwise calibration and identify more adequate model for explanation. Thus, other three-way interaction terms were added respectively in Model 3, 4, and 5 (Table 8). Even though these models reduced the deviance of the model 2, it does not imply that these models improved the model performance, because they did not improve Model 2 significantly considering the change of degree of freedoms. Moreover, the interpretation of them are not more concise than Model 2. Thus, Model 2 was selected as final model. The coefficients of this model is represented in Appendix 1-A.

Table 8. Log-linear model for holiday effect

	Model 1	Model 2	Model 3	Model 4	Model 5
Call	(S) (A) (M) (H)	(S,A) (S,M) (S,H) (A,M) (A,H) (M,H) (S,A,M)	Model 2 + (S,A,H)	Model 2 + (S,M,H)	Model 2 + (S,M,H)
Deviance	2214	28.2	26.6	25.7	16.1
DF	63	34	28	30	22
Pr(>Chisq)	<0.001	0.75	0.54	0.69	0.81
AIC	2088	-39.8	-29.4	-34.3	-27.9

Note: S – Sex / A – age group / M – marital status / H – holiday

Because Model 2 postulates homogeneous associations between holiday effect and the other variables, it is easy to interpret. Compared to the odds of male, the odds of female committing suicide on holidays and post-holidays was 22.1% and 27.2% higher respectively, regardless of age and marital status. It reveals that holiday have greater impact on suicide of female than male. Moreover, regardless of sex and marital status, the odds of middle aged and elderly people was 29.3%, 69.5% higher in post-holidays than the odds of young adults. In examining the relationship between marital status and holiday, the odds of suicide in post-holidays of married people was 35% higher than the odds of the divorced, and the odds of suicide in holiday of the bereaved was 34% higher than the odds of married people, regardless of sex and age. This pattern reveals that the divorced and the bereaved are not vulnerable to postpone effect, while the effect is identified more in older age strata.

4.1.3.3. Copycat effect

The exceptional pattern of temporal variation in 2008 during the study period was found earlier. After the analysis on the daily pattern, ‘copycat effect’ from the suicide of a celebrity, which attracted public attention nationwide, is alleged to be a putative cause of this exceptional pattern. Copycat effect, also known as Werther effect, refers to the phenomenon of suicide committed by celebrity and subsequent surges (Ji et al. 2014). Figure 12 represents daily frequency and cumulative frequency of suicide during 28 days before and after the actress’ suicide. It shows that the famous actress’ suicide played an important role increasing suicides.

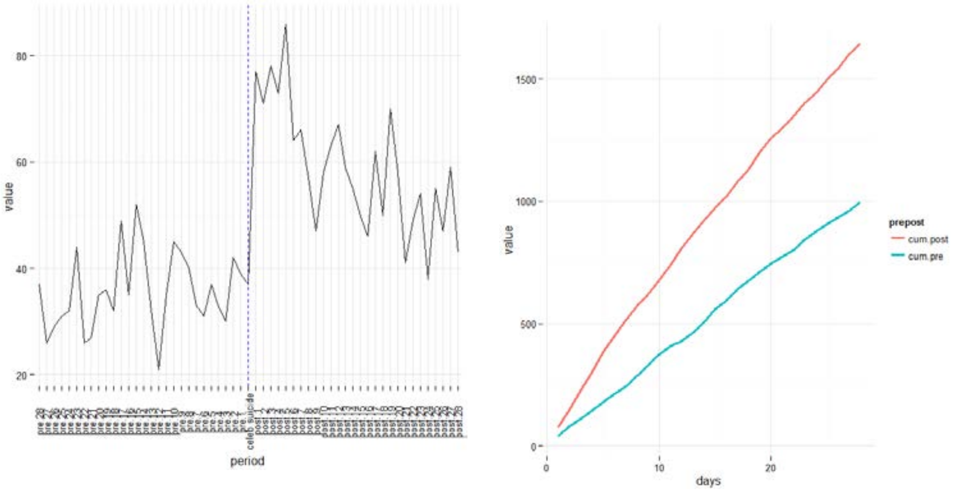


Figure 12. Daily suicide during 28 days before and after the celebrity’s suicide (left); Cumulated suicidal cases for the same period (right)

Previous studies on copycat effect primarily focused on the effect of media coverage (Niederkrotenthaler et al. 2009, Chen et al. 2010, Chen et al. 2012a, Ji et al. 2014). It was revealed that copycat effect significantly affect younger generation, and the effect is relatively weak in older generations (Yip et al. 2006, Niederkrotenthaler et al. 2009). Some studies have found that celebrity’s suicide result in positive effect more on female (Song, 2011, Chen et al. 2012a, Ji et al. 2014), while other studies report that males are more susceptible to media (Chen et al. 2010). Furthermore, the number of suicide committed with the same method used in celebrity’s suicide tends to increase after the suicide event (World Health Organization 2008, Ladwig et al. 2012). This is the reason why World Health Organization recommends not to explicitly report detailed information of method and place of suicide event on media (World Health Organization 2008).

To reveal the discriminative imitating effect of celebrity's suicide in Korea, log-linear model with the variables of sex, age, use of same method (hang in this case), and prior or after the suicide event was developed. Reference factor level used in developing log-linear model was young male who used method different from the celebrity before the celebrity's suicide. Model 1 represents mutual independence model (Table 9). The deviance is very large and the model did not fit well. Compared with mutual independent model, homogeneous association model was calibrated (Model 2). Despite reduction of deviance, fitting is not good yet. Even though three-way interaction term is not easy to interpret, it should be included in the model for fitting data values. Model 3 includes all possible three-way interaction terms, which is the most complex model except saturated model. The final model fits the best with lower AIC, and enables to interpret in more succinct manner; thus the model was selected. Coefficients of the final model are listed in Appendix 1-B.

Table 9. Log-linear model for copycat effect

	Model 1	Model 2	Model 3	Model 4
Call	(S) (A) (M) (C)	(S,A) (S,M) (S,C) (A,M), (A,C), (M,C)	(S,A,M) (S,A,C) (S,M,C) (A,M,C)	(S,A,M) (S,M,C)
Deviance	431	53	5.73	12
DF	25	13	3	12
Pr(>Chisq)	<0.001	<0.001	0.126	0.448
AIC	381	27	−0.271	−12

Note: S – sex / A – age group / M – suicide method (hang / other methods) / C : prior and after the suicide of celebrity

In examining the relationship between sex and celebrity's suicide, it was found that copycat effect in selecting suicidal method was extremely significant. In the case of male, the odds of committing suicide by hang after the celebrity's suicide was 2.113 times higher than the odds of suicide by hang prior to the suicide event. Female was more vulnerable to the imitating effect than male: the odds of female imitating suicide method was 3.755 times higher than that of committing suicide ahead of the celebrity's suicide. As shown in the result, age is independent with celebrity's suicide. More specifically, age is 'conditionally independent' with the suicide event, considering third variables. It shows that the copycat effect was extensively present in all age strata nearly in common.

4.2. Geographical Distribution of Suicide Risks

4.2.1. Standardization of Suicide Rates

From the explorative analysis, it was found that age-sex specific suicide rates are different each other and they change in time. It implies that it will be more diverse in small scales of regions. In addition to this, population structure are different in by region and time. To remove misleading of true geographic distribution of suicidal risks, suicide rates of each region were standardized and mapped, considering properties of suicide cases and population structure in the regions.

4.2.1.1. Directly Standardized Suicide Rates

Age-sex standardized suicide rates were calculated with direct standardization and visualized with choropleth maps to compare suicide risk among regions. Population in 2010 was used as standard population in the calculation. Quantile method was used in classification, because it is the best to guarantee comparability in regions (Brewer and Pickle 2002), which conforms to the mapping objective of DSR maps. Directly standardized rates were classified into 7 quantiles. Suicide rate have unidirectional meaning itself: higher the value, greater the risk. To reflect this characteristic of data value, sequential color scheme, orange to red, was used.

Identifying spatial pattern of metropolitan areas is crucial because large proportion of populations reside in and high−numbers of suicides are recorded in those regions. However, the spatial distribution of suicide risks in metropolitan area might not be seen well due to the nationwide spatial range. Thus, metropolitan areas, including Seoul, Daejeon, Gwangju, Daegu, Ulsan, and Busan were represented together with the nationwide map. Here, DSR map in 2008 is only presented as an example (Fig. 13). The DSR maps in the other years in study period will be represented in appendix 2.

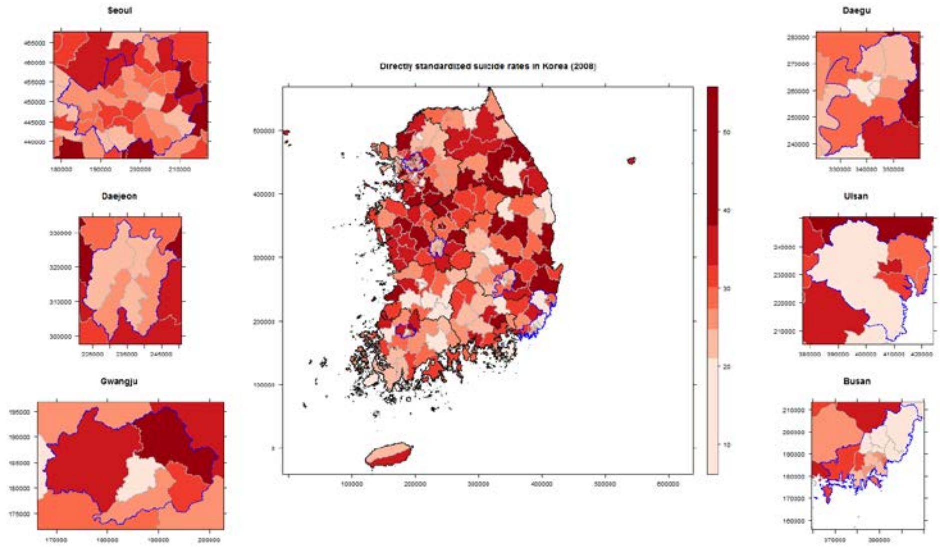


Figure 13. Direct standardized rates of suicide in Seoul, Daejeon, Gwangju, Daegu, Ulsan, Busan, and in Korea (2008)

Based on the comparability of DSR, we can directly compare suicide risks of regions with these map. For example, it can be explained that suicide risks of Western areas in Busan were higher than Eastern side in 2008. In national scale, it seems that suicide risks in Gangwon, the Northeastern part of Korea, and Chungnam, the Western part of Korea, are higher than other areas in national scale. Moreover, suicide risks in Southern and Eastern part of Gyunggi were relatively higher. In contrast, Jeonnam and Jeonbuk, located in Southwestern area, generally represent lower risks. Metropolitan areas commonly show lower risks than the other cities and rural areas, but Gwangju shows relatively high suicide risks among them.

The summary statistics of DSR is shown in Table 10. It shows that mean and median values of DSR in 2009 surged compared to those in 2008, recording the highest suicide risks during the study period. The average DSR decreased slightly in 2010, remained in similar level in 2011, and decreased considerably in 2012, recording lower level of suicide risks than in 2008. Note that minimum value in 2009 is zero, because suicide did not occurred at all in Ulleung-gun in that year.

Table 10. Summary statistics of directly standardized suicide rates (per 100,000)

	2008	2009	2010	2011	2012
Min.	6.2	0.0	14.3	9.9	9.8
1st Qu.	23.9	28.4	28.2	27.5	23.7
Median	**28.1**	**33.0**	**31.9**	**31.5**	**27.6**
Mean	**29.4**	**34.3**	**33.5**	**33.6**	**29.1**
3st Qu.	33.9	38.7	37.8	38.4	33.4
Max	55.6	77.5	62.8	60.1	62.5
Std.dev	8.436	9.432	8.27	8.605	7.891

Because DSR is computed assuming common standard population structure, it is possible to compare longitudinally as well as spatially. To directly compare different regions in different times, choropleth maps should be made using common legend. Note that the ranges of standardized values were different each other in the DSR maps and table 10. If classifications are determined in each time slot, the maps cannot be compared directly because same color represent different intervals of values. To address this problem, entire values in different times should be put together and classified at once. For these purpose, DSR values of the 5-year study period were aggregated and classified to ensure spatio-temporal comparability (Figure 14).

The high-risk areas more concentrated in 2009 nationwide in the map, which corresponds to the result in Table 10. This plot delivers the information on spatio-temporal variation of suicide risks. Thus, temporal changes as well as spatial distribution can be directly interpreted from the map. For example, the suicide risk of Sacheonsi in 2012 (DSR=44.3) can be directly compared to that of Osan-si in 2008 (DSR=19.5) from the map. This map also clearly shows higher risks in Eastern mountainous area and Western plain area, and some parts of Gyunggi. Moreover, lower risks are primarily distributed in Southern part of the country.

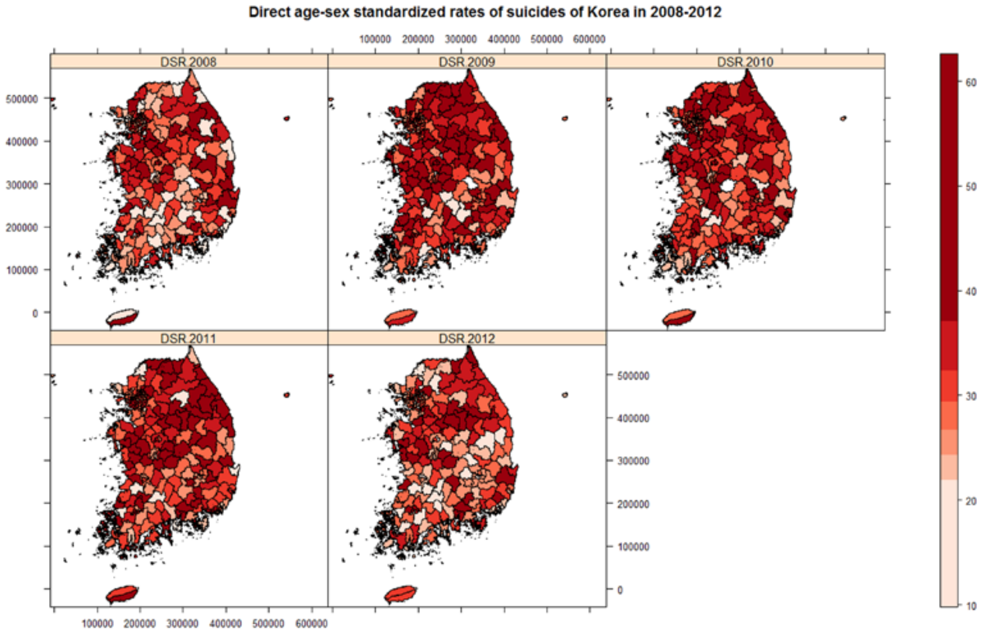


Figure 14. DSR of suicide in Korea (2008-2012)

For summarizing the geographical distribution of suicide risks, the 20 highest and lowest regions were counted by each Si-Do for the study period (Figure 15). As shown in the maps, the top 20 highest suicide risk regions were mainly distributed in Gyunggi, Gangwon, and Chungbuk, while Busan recorded highest cumulative frequency of SGGs with lower suicide risks. Next, the 20 lowest regions were distributed in Southern parts, Jeonnam, Jeonbuk, Gyungnam, and Gyungbuk. Many SGGs in Gyunggi were also included in the top 20 lowest suicide risk in Gyunggi. This implies large geographical variability of risk in Gyunggi. In metropolitan areas, Except Gwangju, cumulative frequency of top 20 lowest suicide risk areas were higher than that of top 20 highest suicide risk areas.

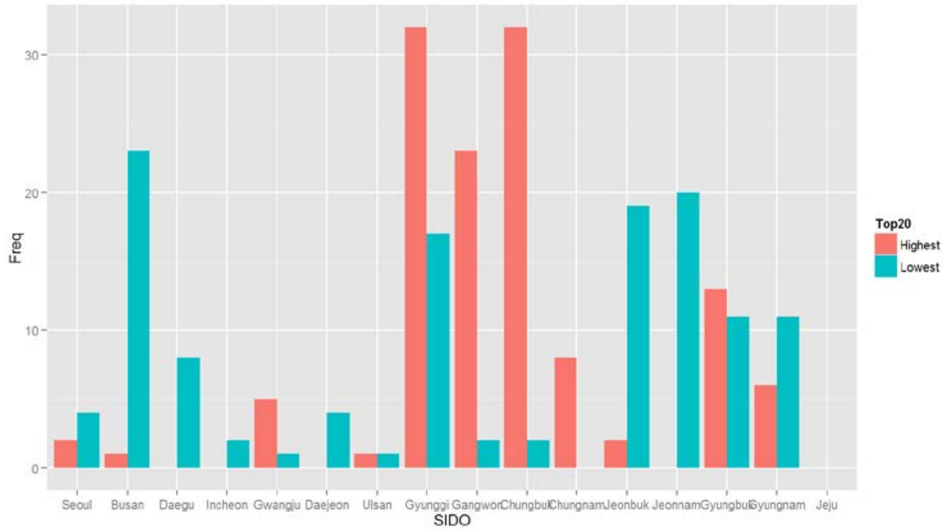


Figure 15. Cumulative barplot of 20 highest and lowest regions, counted by Si-Dos for the study period

4.2.1.2. SMR of Suicide

SMR is commonly used method in disease mapping with indirectly standardization. The purpose of SMR mapping is not in comparability among regions, rather in the suicide burden in each region. SMR was calculated for each year with internal standardization, which means that age-sex specific suicide rates in the total population of each year was used as standard population in the calculation.

In SMR, one is a meaningful value, representing no deficit or excess of observed value relative to expected value, and regarded as a reference value. However, classification by equal intervals in original scale of SMR leads to asymmetric distribution of data values. It comes from the characteristics of ratio, which compresses the data values smaller than the reference value into the range of [0, 1], while the ranges of the values higher than the reference value is unity to infinite. This can be addressed by transformation to log scale, which makes the ranges of both side symmetric, with reference value of $log\,1 = 0$.

For mapping SMR, the color scheme and classification of values were determined following previously performed studies (Middleton et al. 2008, Chang et al. 2011). Red to blue divergent color scheme was used. The intervals of data values were (0−0.5), (0.5−0.67), (0.67−0.9), (0.9, 1.1), (1.1−1.5), (1.5−2) and (>2). This classification is intended to make equal interval in logarithm scale, which is re−transformed to original scale. With this external classification, the maps were made, and the results are shown in Figure 16. SMR maps produced for each year are represented in Appendix 2−B. It should be noted that SMR of different regions cannot be compared and SMR maps in different years as well, even though they share a common legend. In other words, they should only be interpreted relative to the suicide rates in standard population. The geographical pattern of health burdens will be described after stabilizing their statistical variability in the next section.

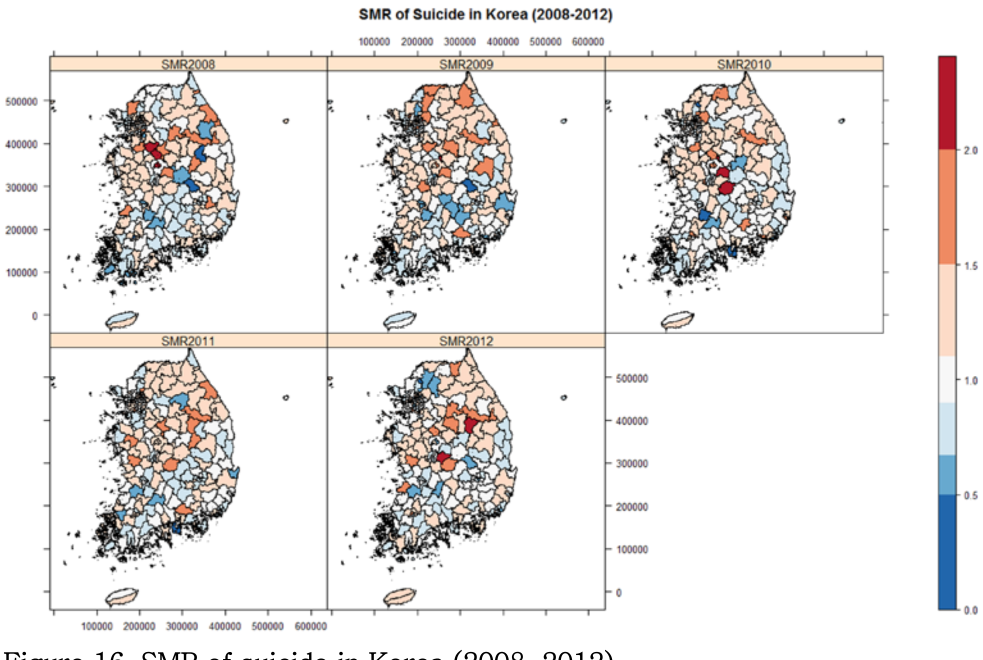


Figure 16. SMR of suicide in Korea (2008−2012)

4.2.2. Bayesian Smoothed Relative Risk

Despite the easy interpretability of SMR, it is often criticized by its statistical instability. As an example, a confidence interval plot of SMR in 2008 is represented in Figure 17. It clearly shows large variability of SMR.

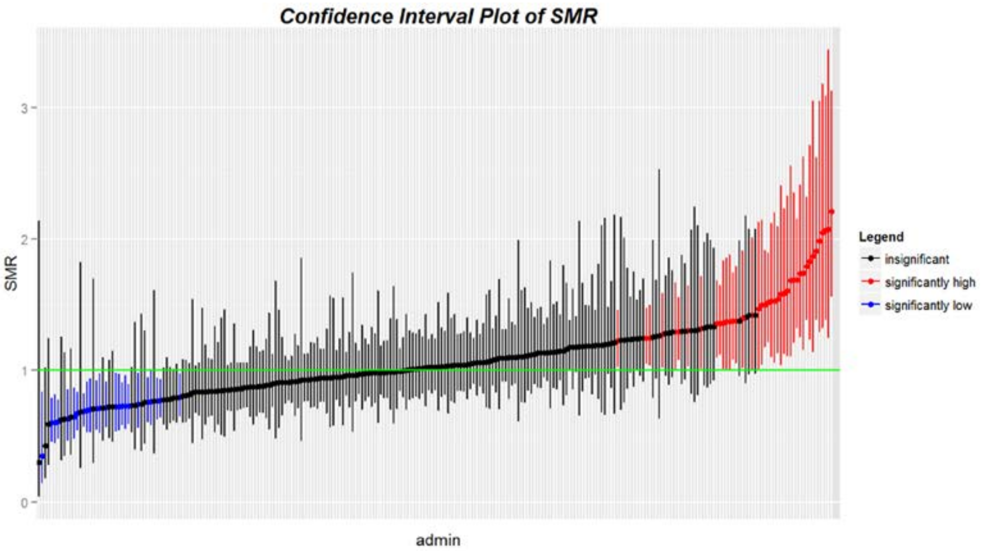


Figure 17. Confidence interval plot of SMR in 2008

To address this problem, RRs of suicide was estimated by empirical Bayes and hierarchical Bayesian method. For empirical Bayes, Poisson−gamma and log−linear model was used for global smoothing, and Marshall's local smoothing technique was used. For hierarchical Bayesian model, Poisson−gamma model and BYM model was implemented. For the estimation of full Bayes method, WinBUGS 1.4.3 software was used to estimate posterior distribution of parameters. Posterior distributions of parameter were estimated by Metropolis−Hastings algorithm, which is a Markov Chain Monte Carlo (MCMC) method. Non−informative gamma distribution were specified as prior distribution. In addition, one Markov Chain was used, and the estimation was derived from 10,000 iterations with 1,000 burn−in period. The mean value of posterior distributions are defined as estimates of RRs. For local estimation, Queen contiguity−based spatial weight matrix was defined and utilized.

Figure 18 shows the maps of RR values estimated by SMR, empirical Bayesian smoothing with Poisson–gamma model (E_PG), log–linear model (E_LN), Marshall's local model (E_ML), hierarchical Bayesian estimation with Poisson–gamma model (PG), and conditional autoregressive model (BYM) with the mortality data in 2010. It clearly shows the smoothing effect of Bayesian estimation on the RRs.

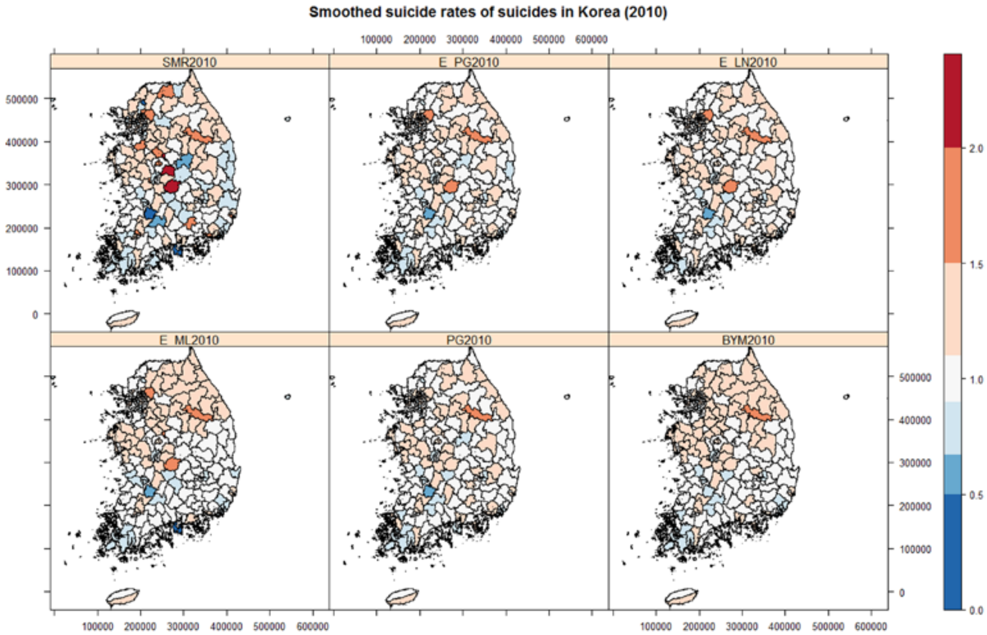


Figure 18. Comparison map of suicide RRs estimated by SMR and Bayesian models in 2010

Box plots of logarithmic transformed SMR and Bayesian smoothed RRs in 2010 (Figure 19), it is lucid that the range of SMR was shrunk when Bayesian smoothing techniques were applied. The plot also shows that hierarchical Bayesian models estimate RRs with lesser variability than empirical Bayesian smoothing techniques.

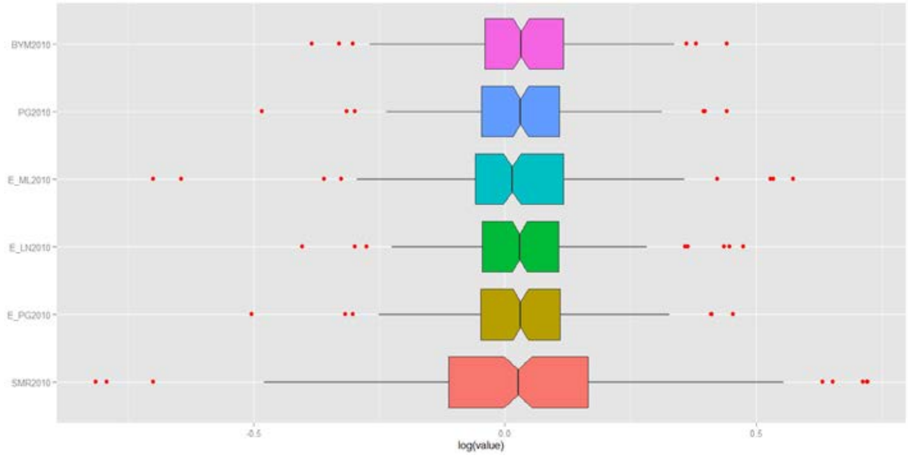


Figure 19. Box plot of relative risks of suicide estimated by SMR and Bayesian smoothing in logarithmic scale (2010)

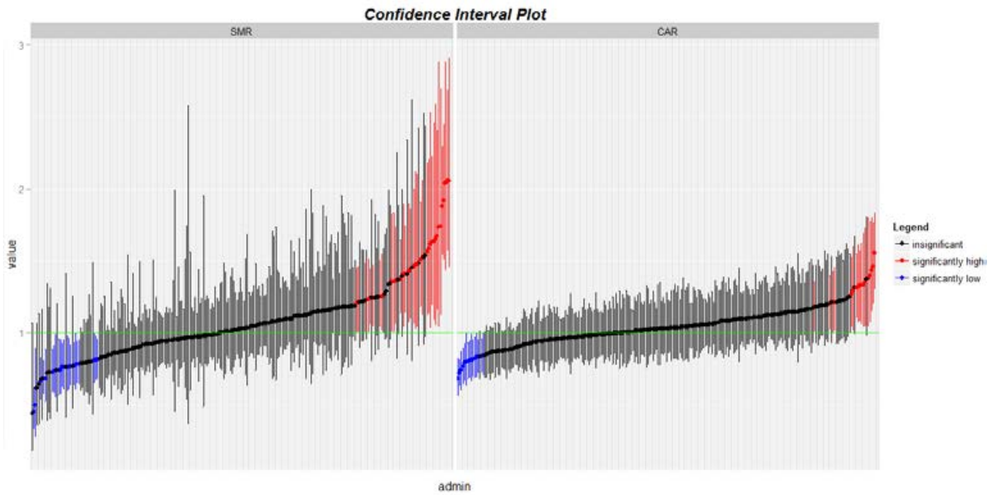


Figure 20. Confidence interval plot of SMR and Hierarchical Bayesian smoothed relative risk (BYM) (2010)

In addition, statistical stability of RRs was enhanced with Bayesian estimation. As an example, confidence interval plots were drawn for SMR values and RRs estimated from BYM model in 2010 (Figure 20). It shows that the Bayesian smoothed estimates are statistically more robust than SMR. As a result, with Bayesian estimation, the variability of RRs was reduced and statistical robustness of estimated values was improved.

The RRs estimated by BYM model are presented here for identifying geographical distribution of RRs (Figure 21). Distinct North–Southern separation have been shown for the study period. The regions in which observed suicide cases outweighed the number of expected values have been mainly distributed from the northeastern to the western areas. In contrast, suicides have been occurred lesser than expected in the southwestern parts.

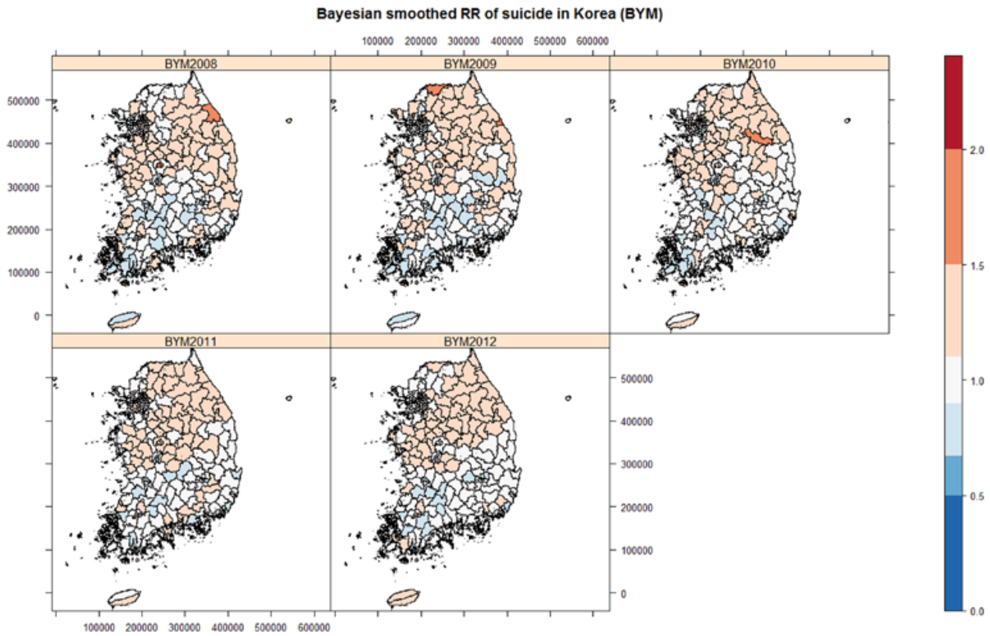


Figure 21. Posterior mean of RRs estimated by BYM model in Korea

In previous studies, “bull’s eye” pattern was found in metropolitan areas. In England, the pattern of increasing suicide rates toward central areas was discovered in metropolitan cities involving London (Middleton et al. 2008). In a study in Australia, similar pattern was found in Western Australia, South Ausralia, Queensland, and New South Wales (Cheung et al. 2012). In a Taiwanese study, the pattern was shown in reverse: inner cities in Taipei represented lower estimates of RRs compared to suburb areas (Chang et al. 2011).

Both patterns were observed in 6 metropolitan areas in Korea. (See Appendix 2–C; the RRs estimated by BYM models were mapped with inset maps of metropolitan areas for each year) From the maps, the bull’s eye pattern is shown in Seoul, Gwangju, and Ulsan. Seoul and Gwangju began to represent the pattern after 2010. In contrast, the reversed pattern has been shown in Daegu and Daejeon consistently during the study period. Busan shows the reversed pattern in 2009, but rather, represents clear distinction between Eastern−Western areas with geographical distribution of lower RRs in Eastern parts of the city.

So far, all estimated RRs were mapped. It means that statistical information was hindered in mapping. The estimated values should be presented in concert with the statistical significance to represent more relevant geographical distribution and to enhance pattern recognition. Thus, following the mapping scheme suggested by Mennis (2006)[12], masking non−significant area, significant RRs were mapped (see Appendix 2−D). By removing insignificant mapped values, statistically more robust geographical patterns were presented. Moreover, compared to the previous maps, spatial patterns were somewhat changed. The bull's eye patterns previously found in Gwangju and Seoul were perished, except Seoul in 2012. Moreover, in Ulsan, the centralized pattern is only remained before 2010. However, the reversed bull's eye patterns represented in Daegu and Daejeon is still remained, and the patterns are clear throughout the study period. In national scale, it is shown that SGGs in Gangwon have taken substantially high suicide burden.

[12] This study discussed efficient approaches for mapping the results of geographically weighted regression. Traditionally, the result of coefficients were illustrated with two plots: parameter estimates and t−value. Three approaches were considered to combine parameter estimates with statistical significance: mapping only signs of coefficient and significance, applying mask to non−significant area, bivariate classification. The first two were suggested as more efficient approaches in the study.

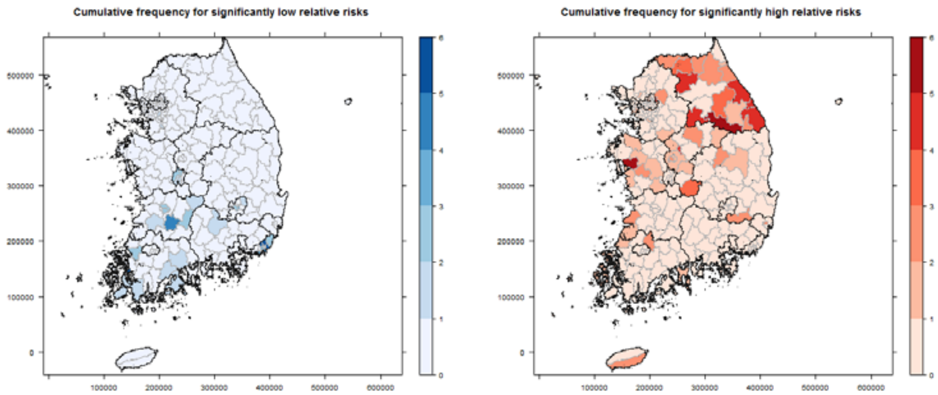


Figure 22. Cumulative frequency for low RR (left); for high RR (right)

To identify temporal variation of statistically robust and significant RRs, the frequencies of the significant values were counted by its magnitudes based on the reference value, unity (Figure 22). As shown in the result, many SGGs in Jeonnam, Jeonbuk, and Busan took less suicide burden during the study period. Five SGGs recorded significantly low RRs for 5 consecutive years: Buk-gu (Busan), Dongnae-gu (Busan), Geumjeong-gu (Busan), Haeundae-gu (Busan), Mokpo-si (Jeonnam). In contrast, Gangwon takes spatially and temporally more suicide burden. Moreover, some regions in Chungnam, Chungbuk, and Gyunggi also shows relatively high RR of suicide. For five years of the study period, Hongseong-gun (Chungnam) and Yeongwol-gun (Gangwon) consistently recorded significantly high RRs.

4.2.3. Cartogram and Web-publishing

4.2.3.1. Cartogram of Suicide

The mapping schemes of rates or RRs derived by standardization and Bayesian smoothing has been discussed and the results were represented so far. It is common to map rates with choropleth maps. However, the information loss of absolute number of cases is inevitable in that case. Cartogram can be an alternative mapping technique to address this problem when it is used as projected base map, providing extra information of interest. Thus, contiguous cartograms were produced, as the number of observed suicide case and the total population were used as the values for density equalizing algorithm.

As an example, the cartograms of directly standardized suicide rates, SMR, and Bayesian smoothed RR in 2010 were made based on the value of absolute number of cases (Figure 23) and population (Figure 24). Note that the color scheme and classifications used in these maps are identical to the results produced earlier. The only change made is that the areas of each region were transformed based on the values used in density-equalizing. Despite the usefulness of cartogram, the difficulty of interpretation is often pointed out as a disadvantage. Not only general map users are not accustomed to the geographical location and shape of regions, but even the practiced map users are not totally aware of the geographical information on the regions. Thus, it is crucial that conventional planimetric map should be included with the cartogram for easy interpretation.

In these cartograms, the area of metropolitan regions, such as Seoul, Busan, and Daegu, expanded extensively in both cartograms. Even though relatively low suicide risks and suicide burdens are observed, it is shown that absolute suicide cases occur more often and greater number of residents live in those areas. In contrast, the shrinkage of Gangwon, for example, implies that suicide cases are less frequently taken place and sparse population reside in each SGG in Gangwon, despite higher suicide risk and RR.

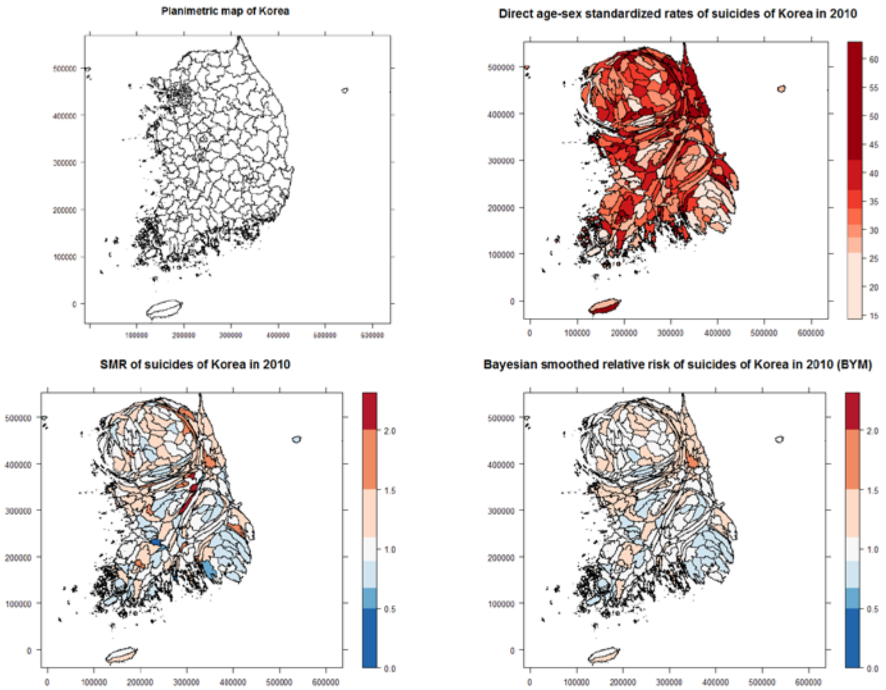


Figure 23. Cartogram of DSR, SMR, and Bayesian smoothed RR in 2010 and planimetric map, based on the number of suicidal cases

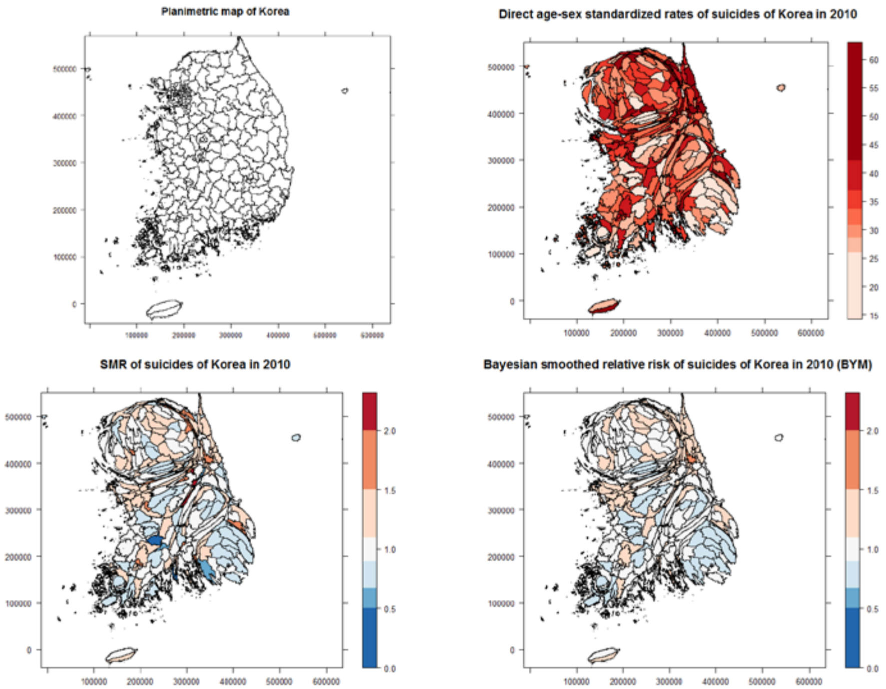


Figure 24. Cartogram of DSR, SMR, and Bayesian smoothed RR in 2010 and planimetric map, based on the population size

4.2.3.2. Interactive Atlas of Suicide

Some of the recent studies in disease mapping urge to publish atlas with web resources as explored in the literature review. It is highly recommended because of its cost-efficiency and the power of impact. Interoperability can be pointed out as a prime virtue of web-mapping. Map users can access and acquire the exact information which they desire to acquire. In this study, the maps were published with the interface of Google Maps and Google Earth. Google Maps has well-forged map data, so that it can be advantageous to visualize geographical information. Moreover, it is easy to web-publish with Google Maps due to its openness with Google API. Google Map API is intended to provide functionality to make interactive maps exploiting the map data of Google. And Google Earth, which use KML file as a standard, can be a useful platform integrating data from different sources (Sun et al., 2012).

All the maps produced in static environment in previous chapters were also produced in HTML format using the R code developed by Kilibarda and Bajat (2012). These files mashup spatial data computed in statistical software over Google Maps. The produced local HTML documents can also be published on web pages. Keyhole Markup Language (KML) files were also produced to visualize the geographical distribution of suicide risks. The R codes developed by Hengl et al. (2013) were utilized for the production of KML files.

As examples, DSR in 2010 and SMR in 2008 were mapped. The interactive maps are useful in some ways. First, it is not needed to design map layout incorporating inset maps to guarantee recognition on the variations in small geographical area, such as Figure 14, because map users are able to zoom-in, zoom-out, and pan as they want. Second, additional information can be provided with map tip function. As shown in the bottom left panel of Figure 25, useful information can be embedded in the map data. While map user could not identify concrete figures of the mapped values in static choropleth maps, for example, now they can check the exact values. Third, statistical map can be compared to the base map overlaid in the background. It is especially advantageous when using cartogram, because planimetric map is not needed to be represented together with cartogram. Lastly, with KML file, time animation maps can be easily produced to visualize spatio-temporal variation of suicide risks. Moreover, the result map can be compared with other KML files which is considered as putative geographical variables, such as temperature.

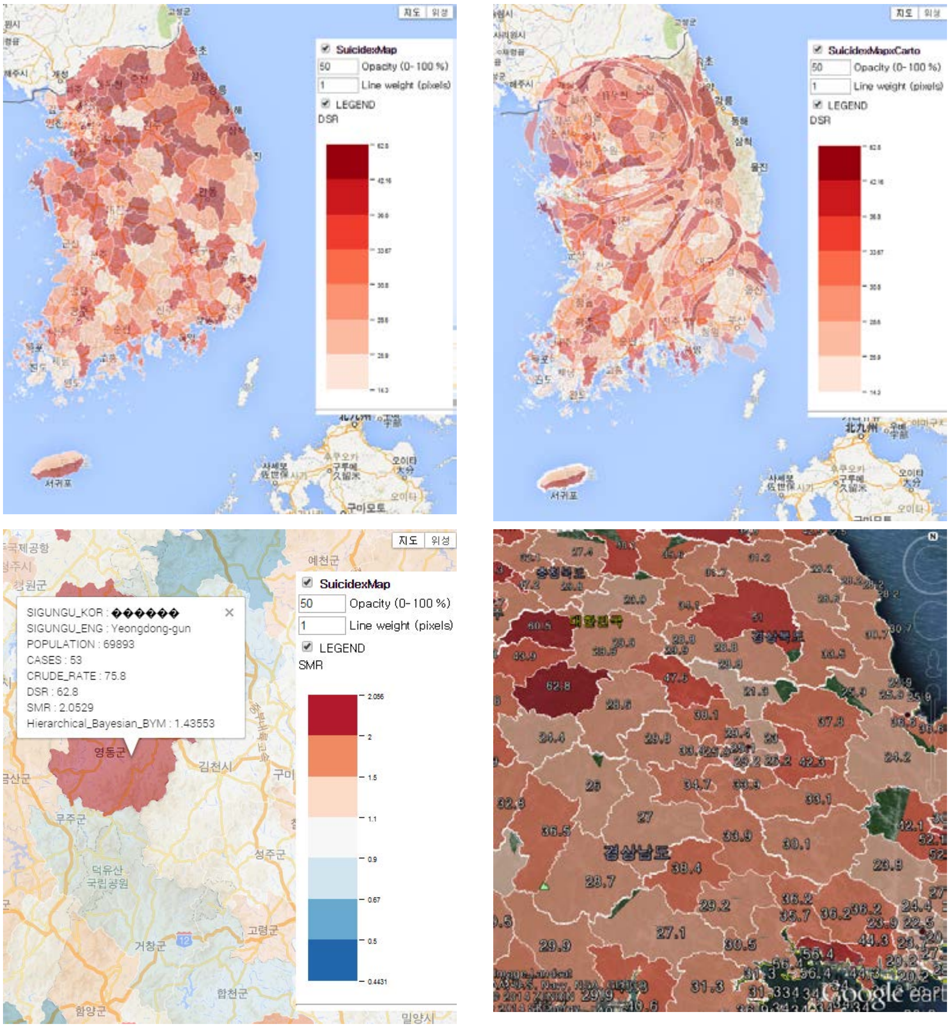


Figure 25. Web mapping with Google Map and Google Earth: DSR of suicide (upper left); cartogram of DSR of suicide (upper right); SMR of suicide, magnified and panned map with map tips (bottom left); KML file of directly standardized suicide rate, represented in Google Earth (bottom right)

4.3. Regional Disparity of Suicide Risks

4.3.1. Aspatial Regional Unevenness of Suicide

Using the directly standardized suicide rates derived in previous section, the extent of the uneven geographical distribution was measured with inequality measures. The indicators summarize the geographical unevenness which were visually identified. To remove the bias which comes from indicator selection, extended Gini index, GEI, D index were used. Moreover, several number of parameters were specified when generalized indices were applied for the purpose of preventing arbitrary selection of parameters.

Regional unevenness of suicide computed by extended Gini index and GEI with different parameters (Figure 26). The longitudinal patterns of both indictors are identical, even when different parameter values were applied in the computation. Regional inequality decreased from 2008 to 2010, but increased continuously after 2011.

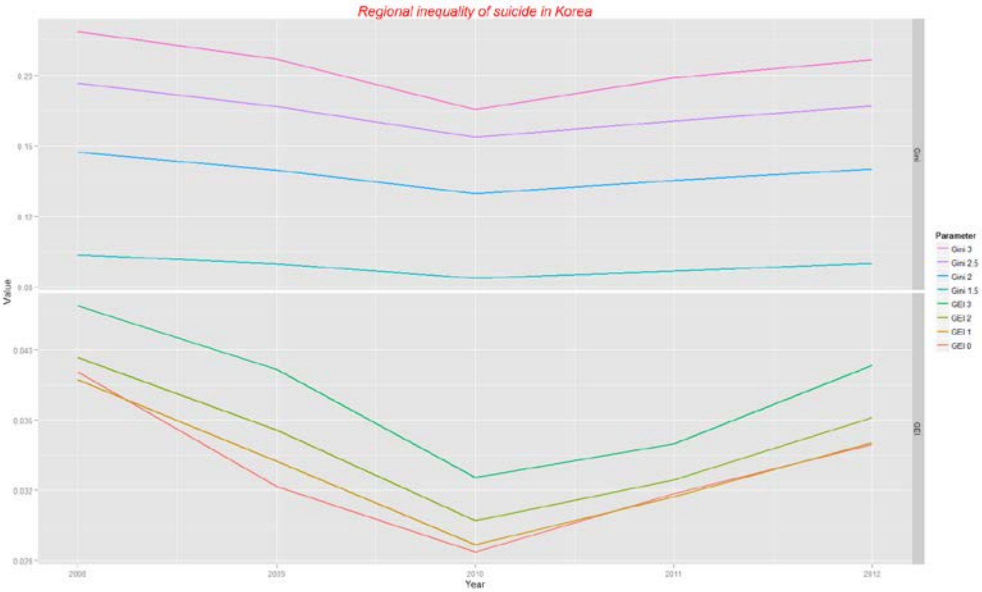


Figure 26. Regional disparity of suicide assessed by extended Gini index and generalized entropy index

The regional inequality was also measured by dissimilarity index (Table 11). It shows the same pattern with the results from extended Gini and GEI. It can be interpreted that about 10 percent of suicide is unequally distributed from the complete evenness.

Table 11. Dissimilarity index (***D***) of suicide

	2008	2009	2010	2011	2012
D	0.1099	0.1022	0.09322	0.1002	0.1044

The variation pattern of inequality can be explained with consideration of higher level of administrative unit, Si–Do. By decomposition of GEI, the contribution of Si–Do on the inequality can be identified. This method decomposes total GEI into and within group inequality; thus the proportions of the inequality among Si–Dos to the total inequality can be derived via decomposition. GEI with parameter 1, which is identical to Theil index, was used in decomposition (Figure 27). As shown in the plot, about a quarters of total inequality comes from the inequality between the administrative units. The proportion of between–Si–Do inequality to the total inequality was 25.46, 29.72, 28.73, 24.26, and 27.53% for each year, taking the largest portion in describing total regional inequality. Except the between–group inequality, regional inequality was largest within Gyunggi. It conforms to the result of Figure 14. It can be used to explain temporal variation of region inequality in Table 11. The decreased inequality between 2008 and 2009 was resulted from the reduced inequality within Gangwon, Gyungbuk, and Gwangju, despite increased inequality within Gyunggi. In 2010, between–Si–Do inequality and within–inequality in Gyunggi decrease substantially, contributing to the temporal variation of total inequality. In 2011, the regional inequality within Gyungnam, and Gyungbuk increased significantly. Total GEI increased despite more reduced within–inequality in Gyunggi. Lastly, in 2012, the increased inequality within Chungbuk and Gyunggi elevated the total inequality, together with increased between–Si–Do inequality.

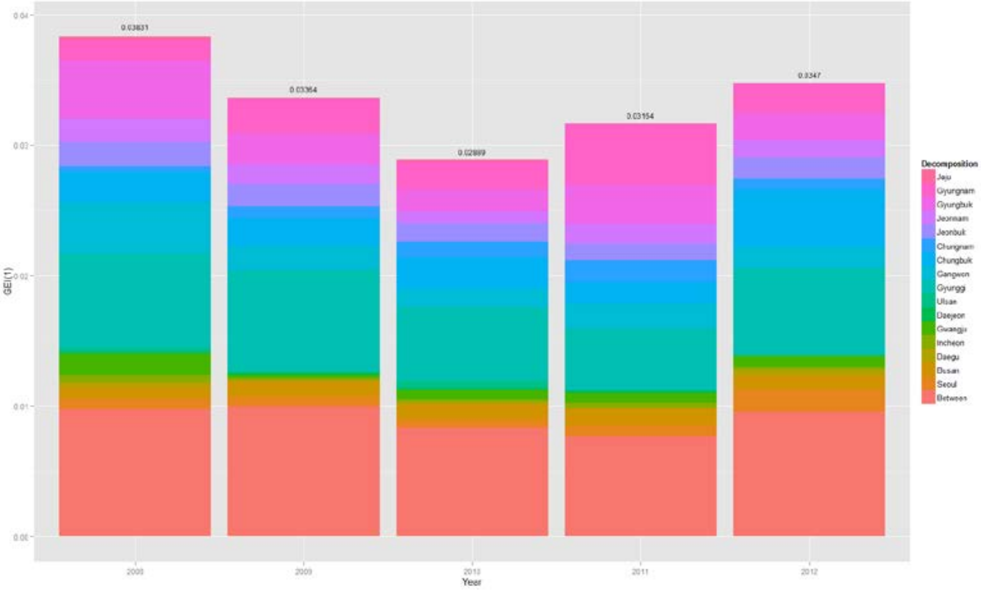


Figure 27. Generalized Entropy Index with parameter 1, decomposed by Si-Do (2008-2012)

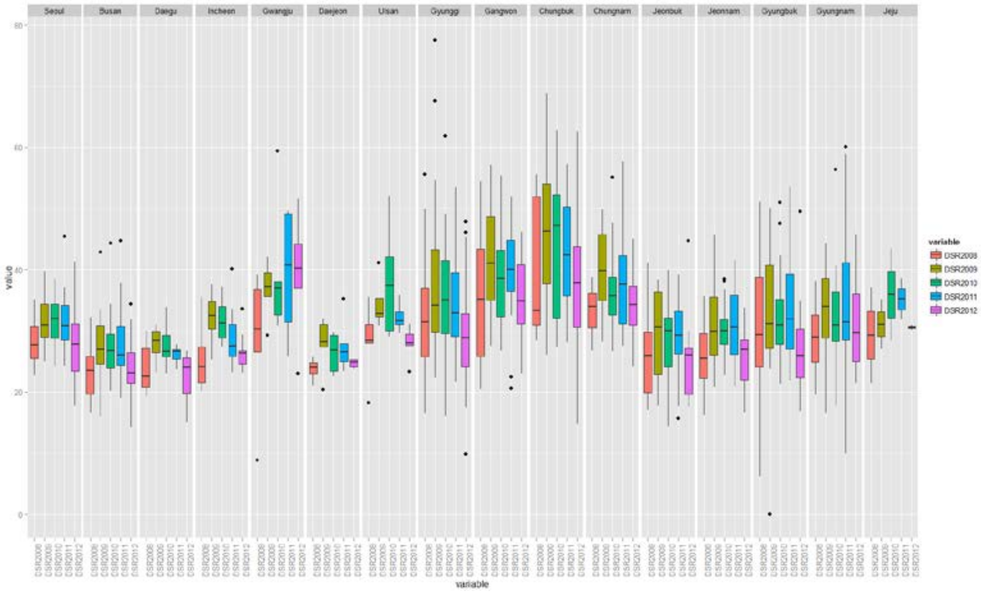


Figure 28. Box plots of DSR, conditioned on each Si-Do and year

Box plots of DSR can also show the temporal variation of inequality, when conditioned on each Si-Do (Figure 28), taking into account of the concept of decomposition within and between Si-Do. The variability of the data distribution within Si-Do can be recognized as within-inequality in GEI decomposition. It is advantageous that this figure also shows the temporal changes of average suicide risks at the same time. Note that, however, it does not provide concrete figure of measured inequality like Figure 27.

In 2008, the suicide risks of metropolitan areas were relatively low, except Gwangju and Ulsan. Suicide risks have increased in almost all Si–Dos in 2009. At that time, the increase rate of Incheon and Gwangju was especially higher than other regions. The range of values in Gangwon, Chungbuk, and Gyungbuk, where the variability had been high in 2008, decreased in 2009. It is supposed that these regions contributed to the diminished regional inequality. In 2010, metropolitan areas except Ulsan recorded slight reduction of suicide rates or so. Other regions where the lower suicide rates were found, such as Jeonnam and Jeonbuk, remained the similar level compared to the previous year. The other regions, except Jeju, either maintained the current level or decreased, contributing to the decreasing regional inequality. In 2011, both suicide risk and range of values in Incheon and Ulsan decreased, while increased in Gwangju. The variance of risks in Gyungnam and Chungnam shows expansion, which might yield increased regional disparity. In 2012, the mean DSR of suicide diminished in all Si–Dos. Especially, suicide rates decreased considerably in Gyungnam. Notwithstanding, the uneven distribution of suicide risks in Chungbuk and Gyunggi contributed on the increase in regional inequality.

The information on the measured regional inequality can be delivered in efficient manner with cumulative frequency legend (Cromley and Cromley 2009). By establishing cumulative frequency legend, it is possible not only to deliver the statistical information of suicide rates, but to merge the information on the regional inequality into the choropleth map. I applied cumulative frequency legend to choropleth mapping for directly standardized rate of suicide (Figure 29). Cumulative curves of suicide case and population were illustrated in the legend, according to the rank order of the rate. When a vertical line is added at the value of grand rate, the distance between two ogives is equal to dissimilarity index,[13] which is equal

[13] The distance is the maximum distance between two ogives.

to the values calculated in Table 11. The bar in x−axis corresponds to the conventional legend, while the bar in y−axis gives statistical information on the distribution of data values.

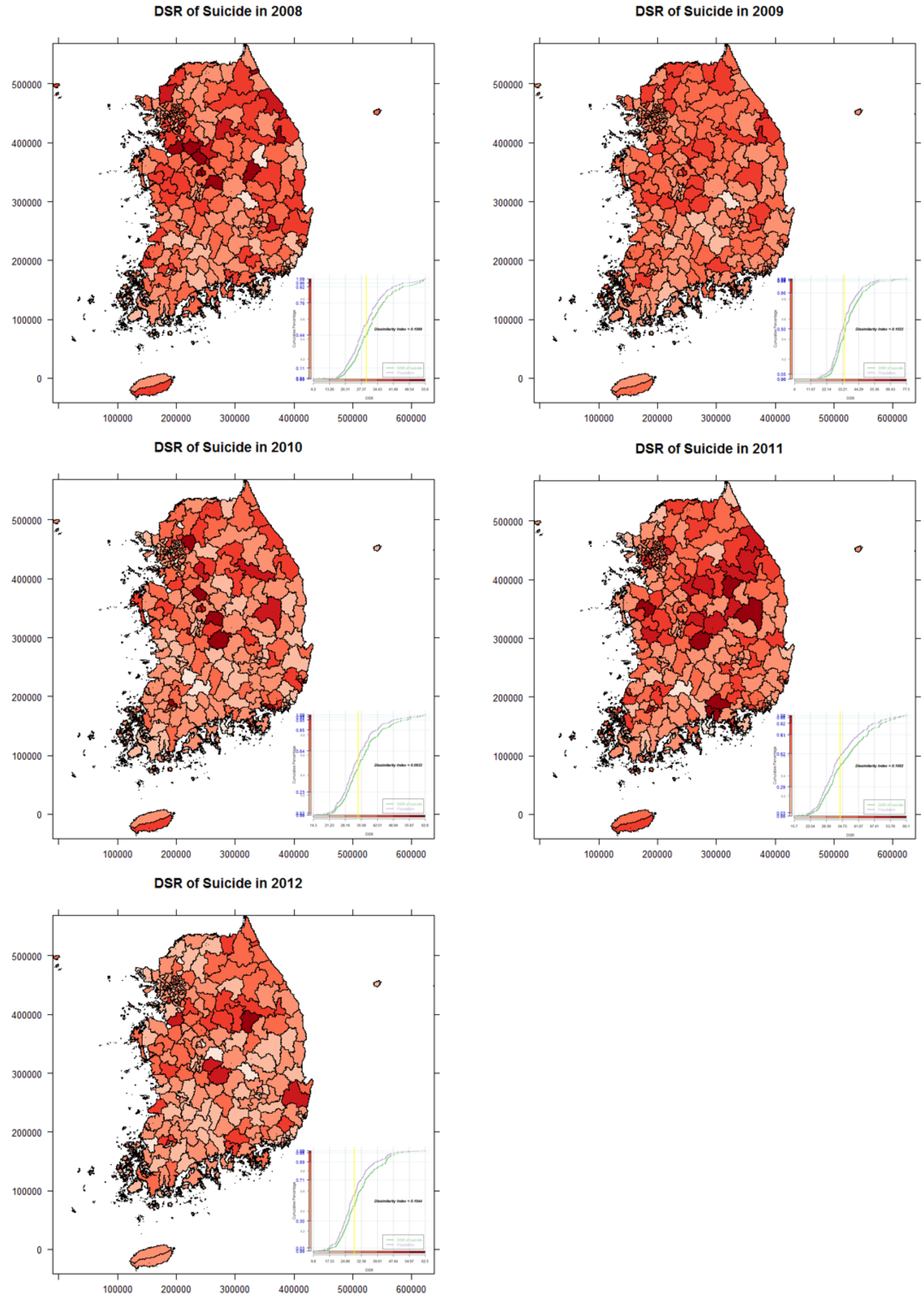


Figure 29. Choropleth map for DSR of suicide with cumulative frequency legend (2008−2012)

4.3.2. Spatial Clustering of Suicide

To explore the spatial characteristics of suicide, the spatial cluster analysis was implemented. First, to identify the presence of spatial dependency, global spatial autocorrelation was measured. In respecting the property of data, I_{cr} was used (Table 12).

Permutation test using Monte Carlo method was performed for 999 times to test the significance of I_{cr}. (Figure 30) The distribution of values derived from the permutation test is represented with histogram and density plot. The blue dotted lines are observed test statistic. It should be noted that p−value was lesser than 0.001 when additional parametric test was implemented. From the result, the null hypothesis of random distribution of suicide can be rejected.

Table 12. Global Spatial Autocorrelation measured by I_{cr}

	2008	2009	2010	2011	2012
Statistic	0.5233	0.5959	0.4649	0.5891	0.5859
p−value	0.001	0.001	0.001	0.001	0.001

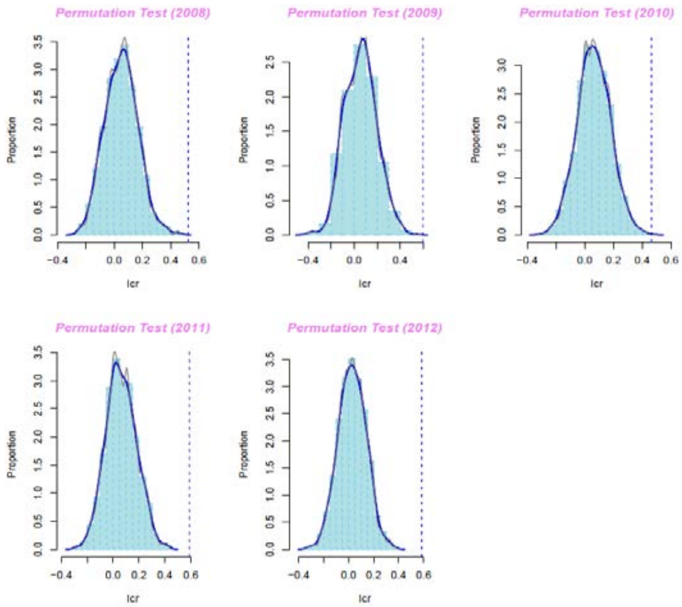


Figure 30. Permutation test of I_{cr} (2008−2012)

Tango's excess events test was also used to diagnose the existence of spatial autocorrelation. Because the result of Tango's spatial clustering test can be changed due to the specification of cluster size (λ), three different values were assigned: The maximum distance between regions following the study of Tango (2000); the distances which were defined as dividing maximum distance by 5 and 10. The tango's EET was tested by negative−binomial parametric model, and the test was replicated 999 times to compute the significance of the observed test statistic. Moreover, the spatial weight proposed by Rogerson (1999) was applied to Tango's EET. And the same cluster sizes were used in the computation. The weights between coordinates regions are plotted in Figure 31 by each clustering size. It gives an information on the difference of weighting schemes, which yields the difference between values of statistics.

The result is shown in Table 13. With significance level of 0.01, it can be concluded that global spatial autocorrelation is shown consistently for the entire study period regardless of cluster size.

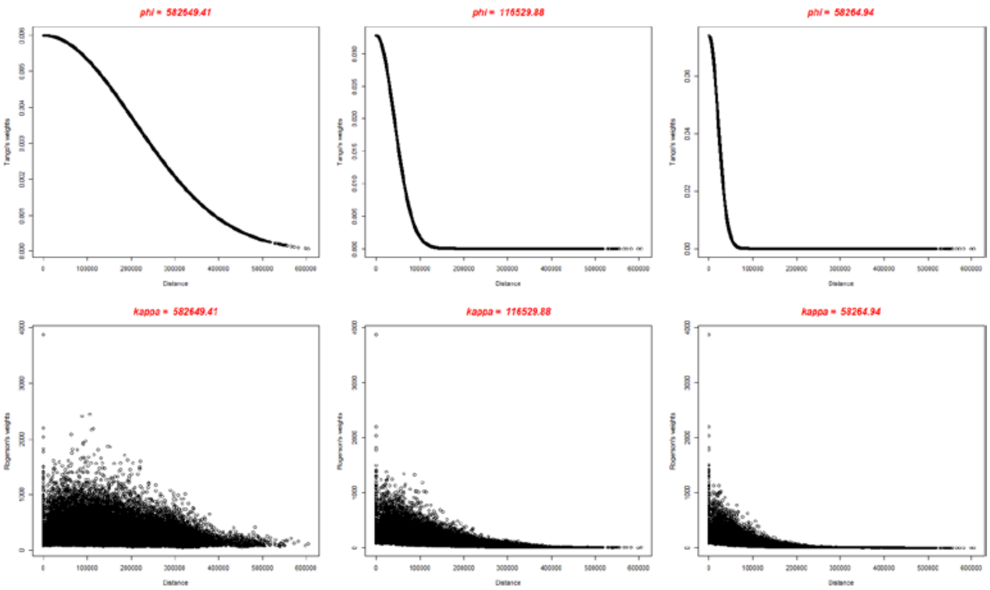


Figure 31. Spatial weights between regions for different cluster sizes

Table 13. Tango's excess events test and Rogerson's test

		Tango's EET		Rogerson's test	
Year	λ	Statistic	p-value	Statistic	p-value
2008	MD	0.000002694	0.022	1.248	<0.001
	MD / 5	0.00002071	0.022	0.8094	<0.001
	MD / 10	0.00004639	0.01	0.5037	<0.001
2009	MD	0.000005559	0.001	1.089	<0.001
	MD / 5	0.00002284	0.006	0.7545	<0.001
	MD / 10	0.00004485	0.013	0.4635	<0.001
2010	MD	0.000003773	0.002	1.091	<0.001
	MD / 5	0.00001862	0.008	0.6946	<0.001
	MD / 10	0.00003824	0.01	0.4248	<0.001
2011	MD	0.000003048	0.002	0.9875	<0.001
	MD / 5	0.00001646	0.018	0.6159	<0.001
	MD / 10	0.00003219	0.011	0.3719	<0.001
2012	MD	0.00003949	0.003	0.8801	<0.001
	MD / 5	0.00002015	0.005	0.5591	<0.001
	MD / 10	0.00003791	0.01	0.344	<0.001

* Note: MD – Maximum distance between the coordinates of regions (=582.6 km)

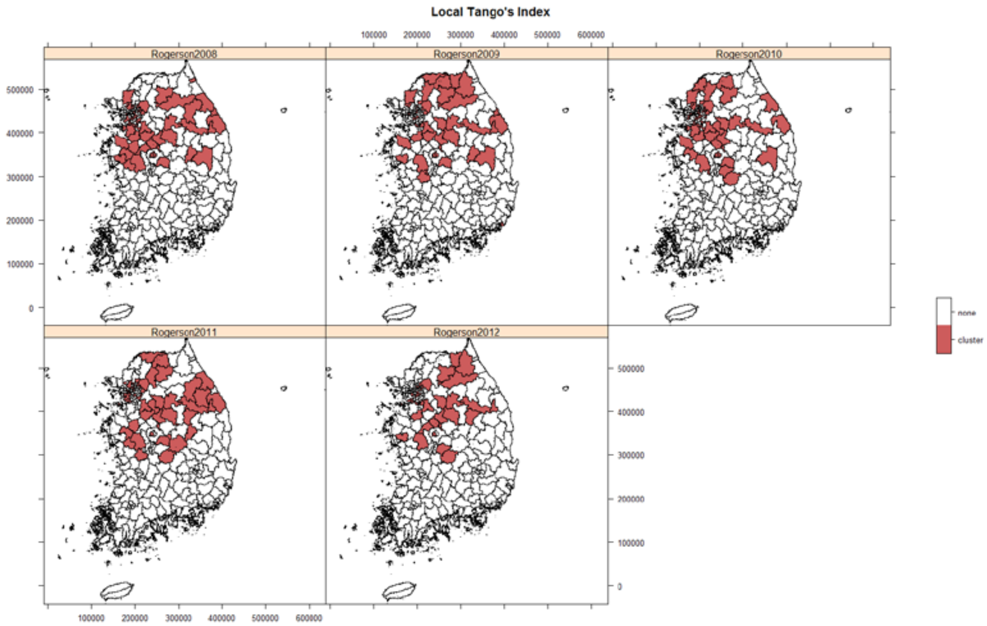


Figure 32. Local spatial chi-square test

Local spatial cluster tests were also implemented to identify where outliers are located in proximately. First, local spatial chi-square goodness-of-fit test was performed to detect local spatial cluster, following Rogerson (1999). For the cluster size parameter (κ), a tenth of the maximum distance between the coordinates of regions was used (Figure 32). It is shown that outliers are spatially distributed in Chungnam, Chungbuk, Gyunggi, and Gangwon.

Local spatial clusters were also identified with local version of I_{cr}. Figure 33 shows I_{cr} scatterplot, which is a modified version of the Moran scatterplot, which was proposed by Anselin (1996). The differences between observed and expected frequency which is standardized by square root of expected number are represented in abscissa, and ordinate is the spatially lagged values of them. With significance level of 0.001, the name of outlier regions are represented in the plot.

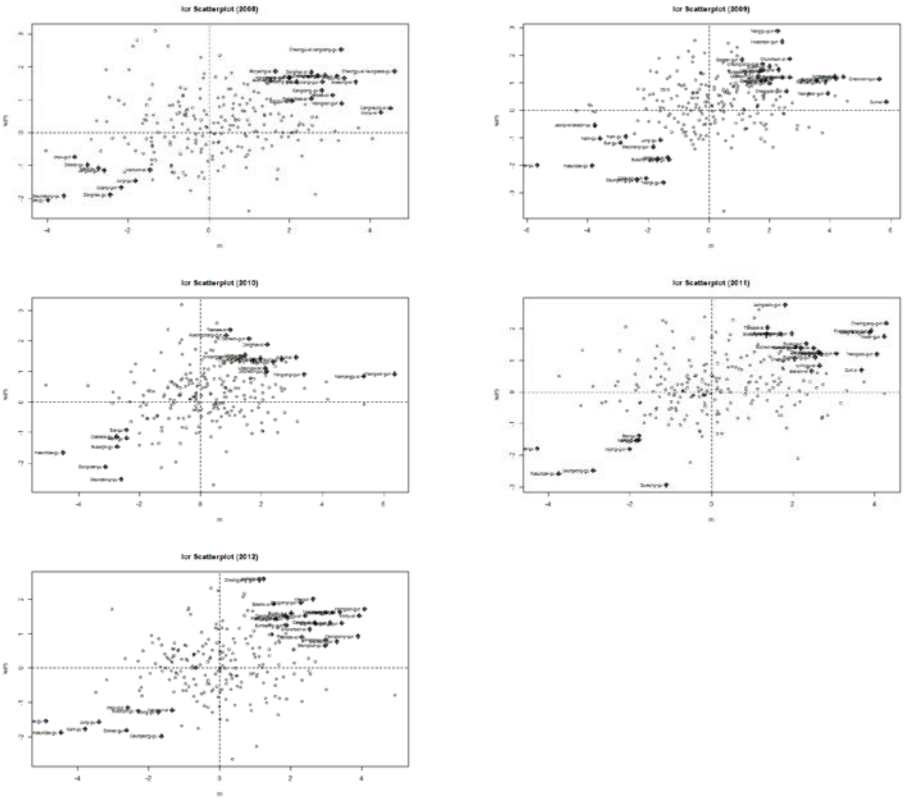


Figure 33. I_{cr} scatterplot (2008-2012)

Based on the plot, significant local clusters were mapped in Figure 34. The region in which the number of observed case is significantly higher than expected value, surrounded by neighboring areas showing similar patterns was defined as 'High–High (HH)' region. It corresponds with the upper–right quadrant in I_{cr} scatterplot. The opposite case was defined as 'Low–Low (LL)', which corresponds with the lower–left quadrant in I_{cr} scatterplot. HH–type spatial clusters are concentrated mainly in Gangwon, Chungnam and Chungbuk, while LL–type spatial clusters are shown in Jeonbuk, Daegu, and Busan (Figure 34). It shows similar spatial pattern with the results of local spatial chi–square test.

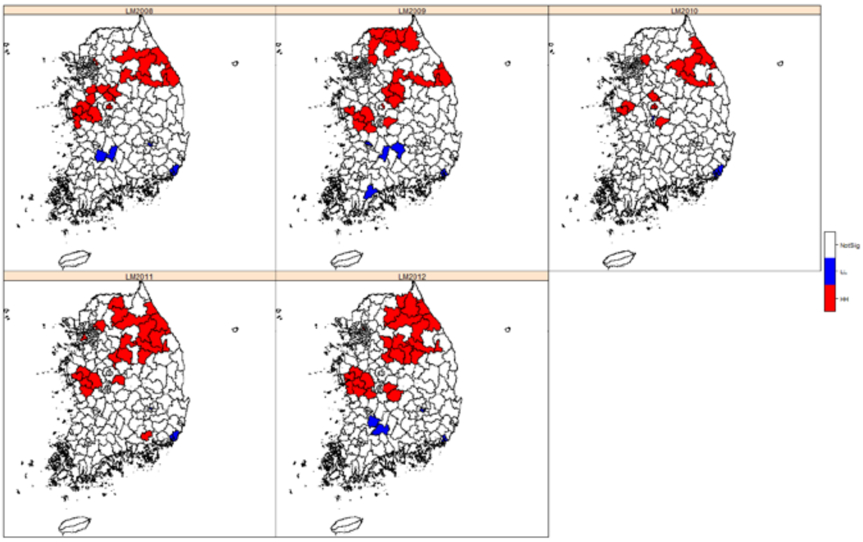


Figure 34. Local cluster of suicide in Korea by Local I_{cr} (2008–2012)

Kulldorff's scan statistic can also be performed to identify the location of spatial cluster (Figure 35). Specifically, this method was applied to find the most likely cluster in this study. The radius of circular window was defined according to the proportion of population inside the window to the total population. In this analysis, the extent of window was determined to contain 5 percent of total population in each window. In figures, MLCs of suicide in each year are represented. The centers of the MLCs were Yeongwol–gun, Chuncheon–si, Hoengseong–gun, Yangyang–gun, and Pyungchang–gun, which are all in Gangwon, in each year.

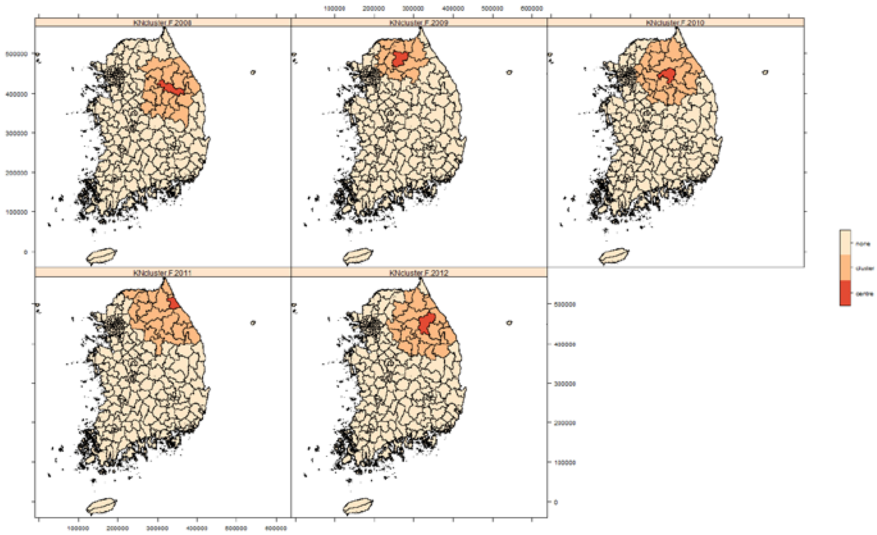


Figure 35. MLC of suicide in Korea scanned by Kulldorff and Nargawalla's method (2008-2012)

4.3.3. Spatial Dissimilarity of Suicide

The analysis of spatial clustering raises the question on the use of aspatial inequality measures, because the spatial autocorrelation is closely related to the regional unevenness. The spatial structure should be considered in the measurement. Thus, the regional inequality of suicide was measured with the spatial dissimilarity indices, and the results are below.

Boundary-modified D index (D_{adj}), proposed by Morrill (1991) was measured first. The result is shown in Table 14. Even after reflecting boundary effect, the pattern of longitudinal change is not changed, compared the result from aspatially measured dissimilarity (Duncan's D).

Table 14. Boundary-modified D_{adj} of suicide

	2008	2009	2010	2011	2012
D_{adj}	0.1074	0.1013	0.09612	0.09703	0.09869

Spatial dissimilarity index ($\breve{D}$) was also calculated following the suggestion Feitosa et al. (2007). First, the local population intensity was computed with Gaussian kernel. To prevent bias from the bandwidth selection, spatial dissimilarity index were computed with different kernel bandwidths. The bandwidths were defined from 0 to 150km by 10km intervals. As shown in Figure 36, temporal pattern is consistent regardless of kernel width. In addition to this, composite population count (equation 36) was computed and applied in the analysis for comparison (Table 15).

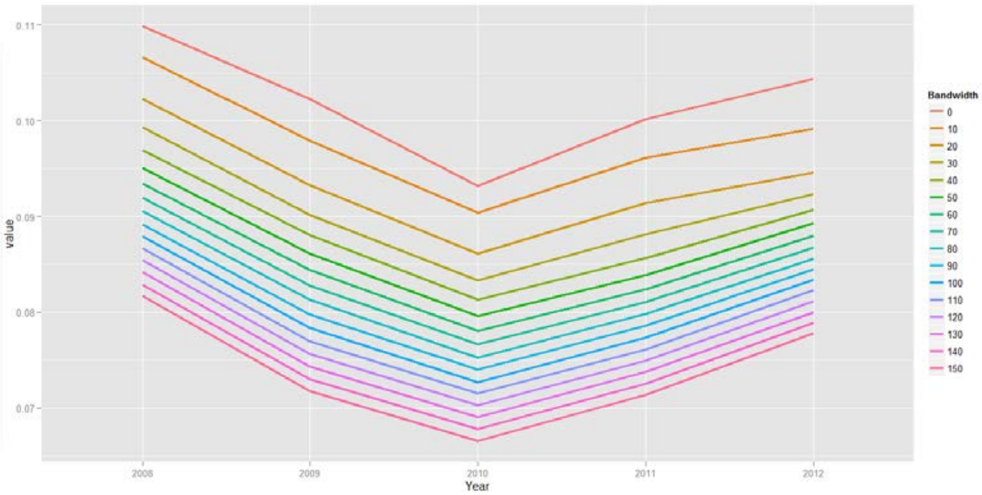


Figure 36. Global spatial dissimilarity indices $\boldsymbol{\breve{D}}$ with different Gaussian kernel bandwidths

Table 15. Global spatial dissimilarity index ($\boldsymbol{\breve{D}}$) of suicide

Local population intensity	Kernel width	2008	2009	2010	2011	2012
Gaussian Kernel	0	0.1099	0.1022	0.09322	0.1002	0.1044
	50	0.0951	0.0861	0.0796	0.0839	0.0893
	100	0.0879	0.0783	0.0727	0.0773	0.0834
	150	0.0817	0.0718	0.0666	0.0713	0.0778
Compositional population count	–	0.06767	0.05824	0.05455	0.0567	0.05882

Regardless of configurations of local population intensity, similar temporal pattern is drawn. The spatial dissimilarity recorded the highest in 2008 and had decreased until 2010. After 2010, the spatial dissimilarity had consistently increased until 2012. It should be noted that the spatial dissimilarity measured by negative exponential kernel shows the identical figures with aspatial D index. It is supposed that the neighboring effect was not effective due to the application of drastic distance decaying function.

The main advantage of the methodology is its extension to the local index $\breve{d}_j$, which comes from the addictiveness in computing global $\breve{D}$. With the local index, the contribution of each region on the total spatial dissimilarity can be mapped and can be easily identified from the map. For the purpose of detecting spatial and temporal variation of local contributions on total dissimilarity, local spatial dissimilarity index was mapped for each year of the study period (Figure 37). Local population intensity was estimated using Gaussian kernel following the original suggestion. The width of the kernel was specified as 50km, which guarantees at least one neighborhood among regions. Youngdong-gun in Chung-buk recorded the highest local dissimilarity index for three years during the study period.

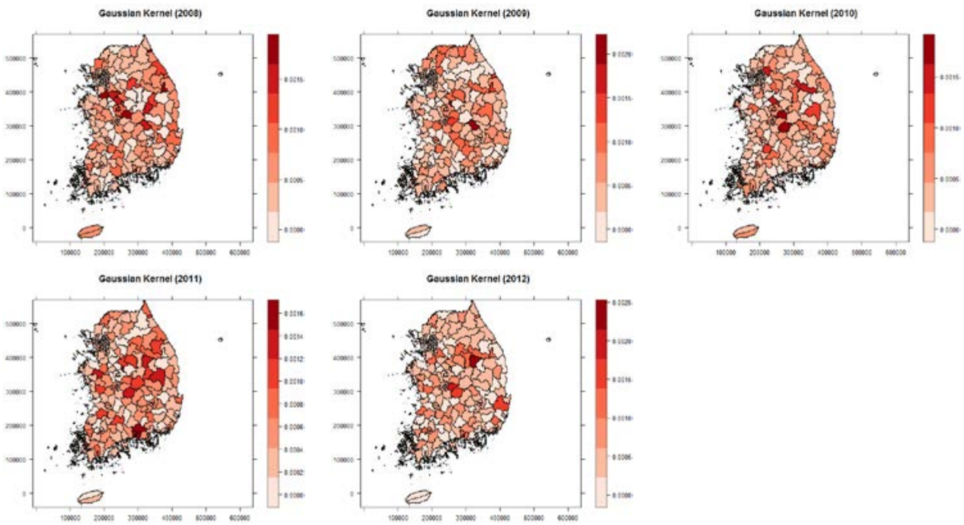


Figure 37. Local spatial dissimilarity index $\breve{d}_j$ for the study period, based on local population intensity using Gaussian kernel

3D mapping can be utilized as an efficient visualization method for representing local spatial dissimilarity index, combining with choropleth map of DSR. The geographical pattern of suicide risks and regional inequality can be represented together exploiting web resources. KML file was produced for DSR in 2008, extruding the height of regions with the values of local of spatial dissimilarity index (Figure 38).

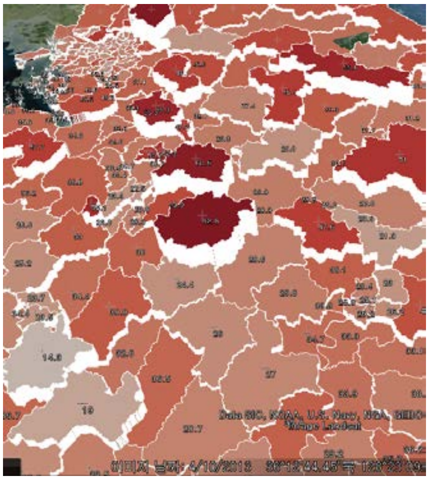


Figure 38. Spatial dissimilarity index represented with vertical height in KML, combined to the choropleth map of DSR of suicide in 2008

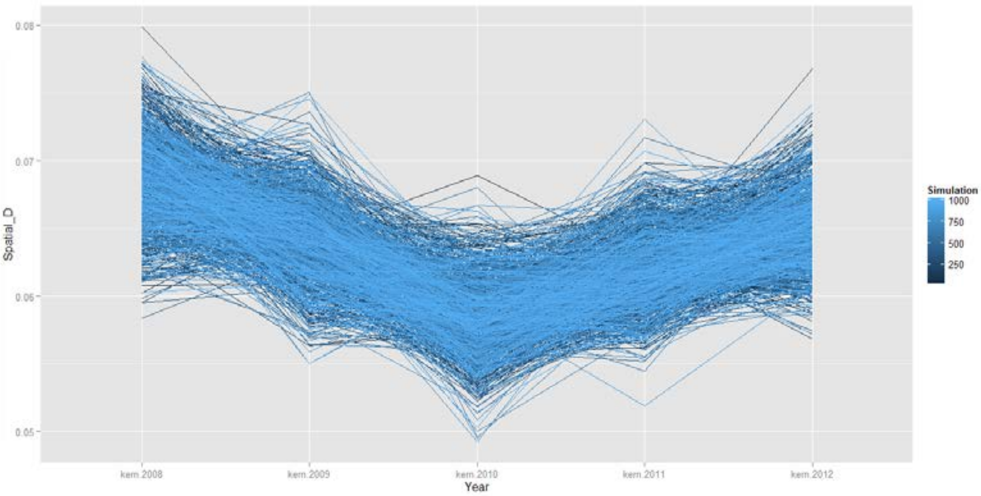


Figure 39. Simulation of random permutation of suicide risks for 1,000 times (Unchanged patterns were shown in 524 times).

It is noteworthy that when spatial dissimilarity was measured, the temporal pattern was not changed compared to that measured with aspatial unevenness. It seems to be caused by stable spatial pattern of suicide risks and local clusters during the study period. In the 1,000 times simulation of randomly permuting suicide risks, different temporal patterns were shown in almost half of simulation cases (Figure 39).

The unchanged pattern might be caused by MAUP as well. Following Reardon and O'Sullivan (2004)'s suggestion, spatial dissimilarity index ($\tilde{D}$) was calculated (Table 15) based on the kernel density surface of suicide case and population. This measure has an advantage that it can resolve Modifiable Areal Unit Problem (MAUP) in some extent. It measures regional dissimilarity based on kernel density of each raster grid, not on the centroids of the areal units. For spatial smoothing, the national spatial range was divided **100 × 100** grids. For each centroid of the grid, kernel density of suicide case and population was calculated with quartic ("biweight") kernel (Figure 40). With the application of equation (33), the spatial dissimilarity index ($\tilde{D}$) was calculated (Table 16).

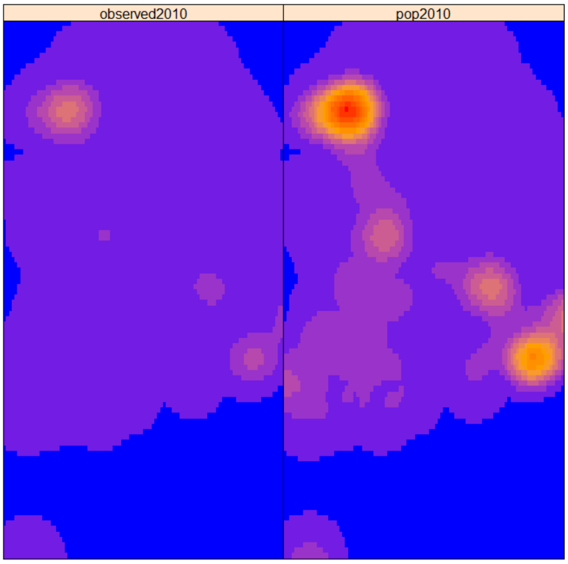


Figure 40. Kernel density surface of observed case and population in 2010

Table 16. Local spatial dissimilarity index ($\widetilde{\boldsymbol{D}}$) of suicide

	2008	2009	2010	2011	2012
$\widetilde{D}$	0.0642	0.0581	0.052	0.0517	0.0547

The result shows different temporal pattern, with comparison to the results of prior measures. In 2011, $\widetilde{D}$ decreased slightly compared to the previous year. It proposes the possibility of the existence of MAUP and need to analysis in smaller scale.

Ⅴ. Conclusion

In recent times, the seriousness of suicide problem has been recognized as a national public health issue. Geographical approach can be an alternative to address the problem. Moreover, health burden and health inequality have gained interest from health practitioners. Thus, this study aimed to explore the geographical pattern of suicide and measure the inequality from regional perspective with national mortality data.

From explorative analysis, it was identified that age-sex specific suicide rates are in different levels from the data. Moreover, the association between categorical variables were examined with log-linear models. The odds of female choosing violent method was higher than male, especially in young adults. With homogeneous association, holiday effect were identified. The odds of elderly people and female were higher than young adults and male, respectively. In addition, female is more vulnerable to the suicide occurrence of celebrity, without regarding to age factors.

This study applied established disease mapping techniques in suicide studies to reveal geographical pattern of suicide in Korea. The map representing DSR of suicide enabled temporal and spatial comparison of suicide risk. Statistically robust RRs were estimated by Bayesian techniques and significant values were mapped. The bull's eye pattern was found in Seoul and Ulsan, while the reversed pattern was shown in Daegu, Daejeon and Busan. Cartogram was used as base projection for visualizing the absolute number of suicide cases, and suicide risks in metropolitan areas cartogram were more highlighted. Lastly, taking into account of advantages of using web-resources, and interactive atlas of suicide risks were produced with the interface of Google Maps and Google Earth.

This study aimed to quantify regional inequality of suicide risks, employing spatial inequality measure, which integrates dimensions of aspatial unevenness and spatial clustering. In the unevenness dimension, utilizing extended Gini index, GEI, and dissimilarity index, the regional inequality were measured, and the regional inequality decreased from 2008 to 2010, but increased in 2011 and 2012. From the decomposition of GEI index, it was found that between-Si-Do inequality took the largest proportions of total inequality. In the spatial clustering dimension, spatial autocorrelation was detected with I_{cr} and spatial chi-square statistics. Moreover, local spatial clusters were searched, and SGGs in Gangwon were selected as the most likely clusters when spatial scan statistics analysis was implemented. The pattern of temporal variation was not changed after taking account of spatial dimension for measuring regional disparity. By simulation, it was found that the stable pattern comes from the invariant spatio-temporal pattern of suicide risks in Korea. Lastly, cumulative frequency legend and 3D mapping was applied for incorporating the information on the regional inequality into the choropleth maps of suicide risks.

Regional inequality of suicide risks is increasing with high spatial correlation, while overall suicide rates is decreasing in recent times. The necessity of region-specific counteracts is implied from the result, and the findings provided by this study might act as explorative bases for making those treatments. For example, suicide prevention center can be established in Chungbuk. Nineteen suicide prevention centers were built nationwide, but not in Chungbuk[14], where high values of suicide risk and local inequality index have been found. Thus, the establishment of a facility in Chungbuk might contribute to reduce the suicide risks of the region and alleviate global regional inequality.

There are several limitations of this study. First, the analysis was based on Si-Gun-Gu, which is relatively large area. It is needed to analyze the spatial variation in more precise level. However, more detailed spatial information on the complete suicide is not permitted to provide to the public, because it is an identifiable and sensitive

[14] Searched from the homepage of the national mental health commission (http://www.nmhc.or.kr/ezboard.php?BID=board21&GID=root&sysop=&fm=&category=3&sido=)

information. Because of the confidentiality, the scale of this analysis was the finest at this time. If detailed information on the smaller-scale area is available in the future, further study could be possible to analyze geographical difference in smaller scale.

Second, time span of this study was limited to recent 5 years. The range of time can be not enough to identify longitudinal change of suicide. Thus, this analysis would be expanded to earlier time period in future studies.

Third, this study focused only on the descriptive and explorative analysis on the mortality data. The spatial pattern and inequality revealed in this study can be modelled for explanation. Future studies would be able to consider various regional characteristics for putative variables, such as socioeconomic status and environmental factors.

Lastly, the methodologies for web-publishing and web-service of suicide atlas, should be developed more exquisitely. It would contribute to the enhancement and the dispersion of geographical knowledge of the public on the health issues in more effective manner

References

Agresti, A. 2002. *Categorical data analysis*. John Wiley & Sons.

---. 2007. *An introduction to categorical data analysis*. John Wiley & Sons.

Ahn, M. H., S. Park, K. Ha, S. H. Choi & J. P. Hong (2012) Gender ratio comparisons of the suicide rates and methods in Korea, Japan, Australia, and the United States. *J Affect Disord,* 142, 161-5.

Ajdacic-Gross, V., M. Bopp, M. Ring, F. Gutzwiller & W. Rossler (2010) Seasonality in suicide--a review and search of new concepts for explaining the heterogeneous phenomena. *Soc Sci Med,* 71, 657-66.

Andres, A. R. (2005) Income inequality, unemployment, and suicide: a panel data analysis of 15 European countries. *Applied Economics,* 37, 439-451.

Andresen, M. A., K. Wuschke, J. B. Kinney, P. J. Brantingham & P. L. Brantingham (2009) Cartograms, crime, and location quotients. *Crime Patterns and Analysis,* 2, 31-46.

Anselin, L. (1995) Local Indicators of Spatial Association - Lisa. *Geographical Analysis,* 27, 93-115.

Anselin, L. (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. *Spatial analytical perspectives on GIS,* 111, 111-125.

Anselin, L., N. Lozano & J. Koschinsky (2006) Rate transformations and smoothing. *Urbana,* 51, 61801.

Assuncao, R. M. & E. A. Reis (1999) A new proposal to adjust Moran's I for population density. *Stat Med,* 18, 2147-62.

Balint, L., P. Dome, G. Daroczi, X. Gonda & Z. Rihmer (2014) Investigation of the marked and long-standing spatial inhomogeneity of the Hungarian suicide rate: A spatial regression approach. *J Affect Disord,* 155, 180-5.

Bando, D. H., A. R. Brunoni, I. M. Benseñor & P. A. Lotufo (2012a) Suicide rates and income in São Paulo and Brazil: a temporal and spatial epidemiologic analysis from 1996 to 2008. *BMC psychiatry,* 12, 127.

Bando, D. H., R. S. Moreira, J. C. Pereira & L. V. Barrozo (2012b) Spatial clusters of suicide in the municipality of São Paulo 1996–2005: an ecological study. *BMC psychiatry,* 12, 124.

Banerjee, S., A. E. Gelfand & B. P. Carlin. 2004. *Hierarchical modeling and analysis for spatial data*. Crc Press.

Besag, J., J. York & A. Mollié (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics,* 43, 1-20.

Best, N., S. Richardson & A. Thomson (2005) A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research,* 14, 35-59.

Beyer, K. M. M., C. Tiwari & G. Rushton (2012) Five Essential Properties of Disease Maps. *Annals of the Association of American Geographers,* 102, 1067-1075.

Braveman, P. (2006) Health disparities and health equity: concepts and

measurement. *Annu Rev Public Health,* 27, 167-94.
Brewer, C. A. & L. Pickle (2002) Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers,* 92, 662-681.
Brown, T., S. McLafferty & G. Moon. 2009. *A companion to health and medical geography*. John Wiley & Sons.
Callanan, V. J. & M. S. Davis (2012) Gender differences in suicide methods. *Social psychiatry and psychiatric epidemiology,* 47, 857-869.
Chang, S. S., D. Gunnell, J. A. Sterne, T. H. Lu & A. T. Cheng (2009) Was the economic crisis 1997-1998 responsible for rising suicide rates in East/Southeast Asia? A time-trend analysis for Japan, Hong Kong, South Korea, Taiwan, Singapore and Thailand. *Soc Sci Med,* 68, 1322-31.
Chang, S. S., J. A. Sterne, B. W. Wheeler, T. H. Lu, J. J. Lin & D. Gunnell (2011) Geography of suicide in Taiwan: spatial patterning and socioeconomic correlates. *Health Place,* 17, 641-50.
Chen, Y.-Y., S.-F. Liao, P.-R. Teng, C.-W. Tsai, H.-F. Fan, W.-C. Lee & A. Cheng (2012a) The impact of media reporting of the suicide of a singer on suicide rates in Taiwan. *Social Psychiatry and Psychiatric Epidemiology,* 47, 215-221.
Chen, Y.-Y., N.-S. Park & T.-H. Lu (2009) Suicide methods used by women in Korea, Sweden, Taiwan and the United States. *Journal of the Formosan Medical Association,* 108, 452-459.
Chen, Y.-Y., P.-C. Tsai, P.-H. Chen, C.-C. Fan, G.-L. Hung & A. A. Cheng (2010) Effect of media reporting of the suicide of a singer in Taiwan: the case of Ivy Li. *Social Psychiatry and Psychiatric Epidemiology,* 45, 363-369.
Chen, Y. Y., K. C. Wu, S. Yousuf & P. S. Yip (2012b) Suicide in Asia: opportunities and challenges. *Epidemiol Rev,* 34, 129-44.
Cheung, Y. T., M. J. Spittal, J. Pirkis & P. S. Yip (2012) Spatial analysis of suicide mortality in Australia: investigation of metropolitan-rural-remote differentials of suicide risk across states/territories. *Soc Sci Med,* 75, 1460-8.
Clayton, D. & J. Kaldor (1987) Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics,* 43, 671-681.
Cowell, F. A. 2000. Measurement of inequality. In *Handbook of income distribution,* eds. A. B. Atkinson & F. Bourguignon, 87-166.
Cromley, R. G. & E. K. Cromley (2009) Choropleth map legend design for visualizing community health disparities. *Int J Health Geogr,* 8, 52.
Denning, D. G., Y. Conwell, D. King & C. Cox (2000) Method choice, intent, and gender in completed suicide. *Suicide and Life-Threatening Behavior,* 30, 282-288.
Dorling, D. & A. Barford (2007) Shaping the world to illustrate inequalities in health. *Bulletin of the World Health Organization,* 85, 890-893.
Dorn, M. L., C. C. Keirns & V. J. Del Casino Jr (2009) Doubting dualisms. *A Companion to Health and Medical Geography*, 55.
Duncan, O. D. & B. Duncan (1955) A Methodological Analysis of Segregation

Indexes. *American Sociological Review,* 20, 210-217.
Durkheim, E. 1952. *Suicide. A Study in Sociology*. London.
Fang, P., S. Dong, J. Xiao, C. Liu, X. Feng & Y. Wang (2010) Regional inequality in health and its determinants: evidence from China. *Health Policy,* 94, 14-25.
Feitosa, F. F., G. Câmara, A. M. V. Monteiro, T. Koschitzki & M. P. S. Silva (2007) Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science,* 21, 299-323.
Gabennesch, H. (1988) When promises fail: A theory of temporal fluctuations in suicide. *Social Forces,* 67, 129-145.
Gastner, M. T. & M. E. Newman (2004) Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 7499-7504.
Gunnell, D., B. Wheeler, S. Chang, B. Thomas, J. Sterne & D. Dorling (2012) Changes in the geography of suicide in young men: England and Wales 1981–2005. *Journal of epidemiology and community health,* 66, 536-543.
Han, S. (2008) Social integration and suicide: The effects of holidays on a decrease in suicide. *Korea journal of population studies,* 31, 165-194. (in Korean)
Hengl, T., P. Roudier, D. Beaudette & E. Pebesma (2013) plotKML: Scientific Visualization of Spatio-temporal Data. *R package*.
Hennig, B. D. 2011. Rediscovering the World: Gridded Cartograms of Human and Physical Space. In *Department of Geography*. University of Sheffield.
Holstein, B. E., C. Currie, W. Boyce, M. T. Damsgaard, I. Gobina, G. Kokonyei, J. Hetland, M. de Looze, M. Richter, P. Due & H. S. I. F. Group (2009) Socio-economic inequality in multiple health complaints among adolescents: international comparative study in 37 countries. *Int J Public Health,* 54 Suppl 2, 260-70.
Howe, G. M. (1964) A national atlas of disease mortality in the United Kingdom. *Geographical Journal*, 15-22.
Im, J. S., S. H. Choi, D. Hong, H. J. Seo, S. Park & J. P. Hong (2011) Proximal risk factors and suicide methods among suicide completers from national suicide mortality data 2004-2006 in Korea. *Compr Psychiatry,* 52, 231-7.
Jessen, G., B. Jensen, E. Arensman, U. Bib-Brahe, P. Crepet, D. D. Leo, K. Hawton, C. Haring, H. Hjelmeland & K. Michel (1999) Attempted suicide and major public holidays in Europe: findings from the WHO/EURO Multicentre Study on Parasuicide. *Acta Psychiatrica Scandinavica,* 99, 412-418.
Jessen, G. & B. F. Jensen (1999) Postponed suicide death? Suicides around birthdays and major public holidays. *Suicide and Life-Threatening Behavior,* 29, 272-283.
Ji, N. J., W. Y. Lee, M. S. Noh & P. S. Yip (2014) The impact of indiscriminate media coverage of a celebrity suicide on a society with a high suicide rate: Epidemiological findings on copycat suicides from South Korea.

J Affect Disord, 156, 56-61.
Kearns, R. & D. Collins. 2010. Health Geography. In *A Companion to Health and Medical Geography,* eds. T. Brown, S. McLafferty & G. Moon. John Wiley & Sons.
Kearns, R. A. (1993) Place and health: towards a reformed medical geography*. *The Professional Geographer,* 45, 139-147.
Kilibarda, M. & B. Bajat (2012) plotgooglemaps: The r-based web-mapping tool for thematic spatial data. *Geomatica,* 66, 37-49.
Kim, H., Y. J. Song, J. J. Yi, W. J. Chung & C. M. Nam (2004) Changes in mortality after the recent economic crisis in South Korea. *Ann Epidemiol,* 14, 442-6.
Kim, S. Y., M.-H. Kim, I. Kawachi & Y. Cho (2011a) Comparative epidemiology of suicide in South Korea and Japan: effects of age, gender and suicide methods. *Crisis: The Journal of Crisis Intervention and Suicide Prevention,* 32, 5-14.
Kim, Y., H. Kim & D. S. Kim (2011b) Association between daily environmental temperature and suicide mortality in Korea (2001-2005). *Psychiatry Res,* 186, 390-6.
Korea National Statistical Office. 2008. Annual report on the cause of death statistics 2007. Daejeon: Korea National Statistical Office. (in Korean)
---. 2009. Annual report on the cause of death statistics 2008. Daejeon: Korea National Statistical Office. (in Korean)
---. 2010. Annual report on the cause of death statistics 2009. Daejeon: Korea National Statistical Office. (in Korean)
---. 2011. Annual report on the cause of death statistics 2010. Daejeon: Korea National Statistical Office. (in Korean)
---. 2012. Annual report on the cause of death statistics 2011. Daejeon: Korea National Statistical Office. (in Korean)
---. 2013. Annual report on the cause of death statistics 2012. Daejeon: Korea National Statistical Office. (in Korean)
Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics - Theory and Methods,* 26, 1481-1496.
Kulldorff, M. & N. Nagarwalla (1995) Spatial disease clusters: detection and inference. *Stat Med,* 14, 799-810.
Ladwig, K.-H., S. Kunrath, K. Lukaschek & J. Baumert (2012) The railway suicide death of a famous German football player: Impact on the subsequent frequency of railway suicide acts in Germany. *Journal of Affective Disorders,* 136, 194-198.
Legendre, P. & L. F. Legendre. 2012. *Numerical ecology*. Elsevier.
Litchfield, J. A. (1999) Inequality: methods and tools. *Text for the World Bank PovertyNet*.
Lorant, V., A. E. Kunst, M. Huisman, G. Costa & J. Mackenbach (2005) Socio-economic inequalities in suicide: a European comparative study. *The British journal of psychiatry,* 187, 49-54.
Marshall, R. J. (1991) Mapping Disease and Mortality-Rates Using Empirical Bayes Estimators. *Applied Statistics-Journal of the Royal Statistical Society Series C,* 40, 283-294.
Massey, D. S. & N. A. Denton (1988) The Dimensions of Residential

Segregation. *Social Forces,* 67, 281-315.
Mayer, J. D. 2010. Medical Geography. In *A Companion to Health and Medical Geography,* eds. T. Brown, S. McLafferty & G. Moon. John Wiley & Sons.
Mennis, J. (2006) Mapping the Results of Geographically Weighted Regression. *The Cartographic Journal,* 43, 171-179.
Middleton, N., J. A. Sterne & D. J. Gunnell (2008) An atlas of suicide mortality: England and Wales, 1988-1994. *Health Place,* 14, 492-506.
Moran, P. A. P. (1948) The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society Series B-Statistical Methodology,* 10, 243-251.
Morrill, R. L. 1991. On the measure of geographical segregation. In *Geography research forum*, 25-36.
Niederkrotenthaler, T., B. Till, N. D. Kapusta, M. Voracek, K. Dervic & G. Sonneck (2009) Copycat effects after media reports on suicide: a population-based ecologic study. *Soc Sci Med,* 69, 1085-90.
Nishi, M., H. Miyake, H. Okamoto, Y. Goto & T. Sakai (2000) Relationship between suicide and holidays. *Journal of Epidemiology/Japan Epidemiological Association,* 10, 317-320.
Pearce, J., R. Barnett & I. Jones (2007) Have urban/rural inequalities in suicide in New Zealand grown during the period 1980-2001? *Soc Sci Med,* 65, 1807-19.
Peterson, M. P. 2003. *Maps and the Internet*. Elsevier.
Pickle, L. W. (2009) A history and critique of U.S. mortality atlases. *Spat Spatiotemporal Epidemiol,* 1, 3-17.
Qi, X., W. Hu, A. Page & S. Tong (2012) Spatial clusters of suicide in Australia. *BMC Psychiatry,* 12, 86.
Qi, X., S. Tong & W. Hu (2010) Spatial distribution of suicide in Queensland, Australia. *BMC Psychiatry,* 10, 106.
Reardon, S. F. & D. O'Sullivan (2004) Measures of spatial segregation. *Sociological methodology,* 34, 121-162.
Rey, S. J. & R. J. Smith (2013) A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences,* 6, 55-70.
Rezaeian, M., G. Dunn, S. St Leger & L. Appleby (2007) Do hot spots of deprivation predict the rates of suicide within London boroughs? *Health Place,* 13, 886-93.
Richardson, S., A. Thomson, N. Best & P. Elliott (2004) Interpreting Posterior Relative Risk Estimates in Disease-Mapping Studies. *Environmental Health Perspectives,* 112, 1016-1025.
Rogerson, P. A. (1999) The Detection of Clusters Using a Spatial Version of the Chi-Square Goodness-of-Fit Statistic. *Geographical Analysis,* 31, 130-147.
Saunderson, T. R. & I. H. Langford (1996) A study of the geographical distribution of suicide rates in England and Wales 1989-92 using empirical bayes estimates. *Soc Sci Med,* 43, 489-502.
Shaw, M., D. Dorling & R. Mitchell. 2002. *Health, place, and society*. Pearson Education.
Snow, J. 1855. *On the mode of communication of cholera*. London: John

Churchill.
Song, Y. 2011. The comparative study of change in characteristics of suicidal attempts before and after the celebrity suicide. In Department of Medicine, Korea: Yonsei University. (in Korean)
Tango, T. (1984) The detection of disease clustering in time. *Biometrics,* 40, 15-26.
--- (1995) A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat Med,* 14, 2323-34.
Unnithan, N. P. & H. P. Whitt (1992) Inequality, economic development and lethal violence: a cross-national analysis of suicide and homicide. *International Journal of Comparative Sociology,* 33, 3-4.
Wagstaff, A., P. Paci & E. van Doorslaer (1991) On the measurement of inequalities in health. *Soc Sci Med,* 33, 545-57.
Wakefield, J. (2007) Disease mapping and spatial regression with count data. *Biostatistics,* 8, 158-83.
Waller, L. A. & C. A. Gotway. 2004. *Applied spatial statistics for public health data*. John Wiley & Sons.
Walter, S. (1992) The analysis of regional patterns in health data I. Distributional Considerations. *American Journal of Epidemiology,* 136, 730-741.
Wang, F. (2004) Spatial clusters of cancers in Illinois 1986-2000. *J Med Syst,* 28, 237-56.
White, M. J. (1983) The Measurement of Spatial Segregation. *American Journal of Sociology,* 88, 1008-1018.
Wong, D. W. (2002) Spatial Measures of Segregation and Gis. *Urban Geography,* 23, 85-92.
Wong, D. W. (2005) Formulating a general spatial segregation measure. *Professional Geographer,* 57, 285-294.
Wong, D. W. S. (1993) Spatial Indexes of Segregation. *Urban Studies,* 30, 559-572.
World Health Organization. 2008. Preventing Suicide: A Resource for Media Professionals. Geneva: WHO.
---. 2012. *Public health action for the prevention of suicide: a framework*. World Health Organization.
---. 2013. *Handbook on health inequality monitoring with a special focus on low-and middle-income countries*. World Health Organization.
Wu, K. C., Y. Y. Chen & P. S. Yip (2012) Suicide methods in Asia: implications in suicide prevention. *Int J Environ Res Public Health,* 9, 1135-58.
Yip, P. S., Y. Y. Chen, S. Yousuf, C. K. Lee, K. Kawano, V. Routley, B. C. Ben Park, T. Yamauchi, H. Tachimori, A. Clapperton & K. C. Wu (2012) Towards a reassessment of the role of divorce in suicide outcomes: evidence from five Pacific Rim populations. *Soc Sci Med,* 75, 358-66.
Yip, P. S., K.-W. Fu, K. C. Yang, B. Y. Ip, C. L. Chan, E. Y. Chen, D. T. Lee, F. Y. Law & K. Hawton (2006) The effects of a celebrity suicide on suicide rates in Hong Kong. *Journal of affective disorders,* 93, 245-252.
Yitzhaki, S. (1983) On an extension of the Gini inequality index. *International Economic Review*, 617-628.

Appendix

Appendix 1. Coefficients of the log-linear models

1-A. Holiday effect model

	Estimate	Std.Error	z.value	Pr(>|z|)
(Intercept)	4.299	0.094	45.499	<0.001
Holidayholiday	-0.114	0.116	-0.976	0.329
Holidaypost	0.054	0.113	0.476	0.634
Sexfemale	-0.583	0.125	-4.681	<0.001
Age_MiddleAged	0.7	0.108	6.498	<0.001
Age_Elderly	0.339	0.123	2.743	0.006
Marriageunmarried	0.834	0.11	7.576	<0.001
Marriagedivorced	-1.058	0.157	-6.743	<0.001
Marriagebereaved	-3.262	0.355	-9.199	<0.001
Age_MiddleAged:Marriageunmarried	-2.401	0.139	-17.261	<0.001
Age_Elderly:Marriageunmarried	-4.9	0.42	-11.675	<0.001
Age_MiddleAged:Marriagedivorced	0.315	0.159	1.976	0.048
Age_Elderly:Marriagedivorced	-1.044	0.212	-4.932	<0.001
Age_MiddleAged:Marriagebereaved	0.196	0.397	0.493	0.622
Age_Elderly:Marriagebereaved	1.887	0.358	5.27	<0.001
sexfemale:Marriageunmarried	-0.382	0.137	-2.781	0.005
sexfemale:Marriagedivorced	0.127	0.212	0.598	0.55
sexfemale:Marriagebereaved	0.023	0.538	0.042	0.966
sexfemale:Age_MiddleAged	-0.532	0.138	-3.862	<0.001
sexfemale:Age_Elderly	-1.218	0.166	-7.334	<0.001
holidayholiday:Age_MiddleAged	0.108	0.122	0.882	0.378
holidaypost:Age_MiddleAged	0.257	0.119	2.168	0.03
holidayholiday:Age_Elderly	0.081	0.152	0.531	0.596
holidaypost:Age_Elderly	0.528	0.142	3.705	<0.001
holidayholiday:Marriageunmarried	-0.033	0.131	-0.25	0.803
holidaypost:Marriageunmarried	-0.209	0.128	-1.64	0.101
holidayholiday:Marriagedivorced	0.111	0.136	0.815	0.415
holidaypost:Marriagedivorced	-0.301	0.136	-2.214	0.027
holidayholiday:Marriagebereaved	0.297	0.167	1.779	0.075
holidaypost:Marriagebereaved	-0.063	0.158	-0.397	0.692
holidayholiday:sexfemale	0.2	0.103	1.932	0.053
holidaypost:sexfemale	0.24	0.099	2.418	0.016
sexfemale:Age_MiddleAged:Marriageunmarried	-0.573	0.329	-1.743	0.081
sexfemale:Age_Elderly:Marriageunmarried	1.837	0.633	2.9	0.004
sexfemale:Age_MiddleAged:Marriagedivorced	-0.552	0.274	-2.017	0.044
sexfemale:Age_Elderly:Marriagedivorced	-0.863	0.578	-1.493	0.135
sexfemale:Age_MiddleAged:Marriagebereaved	1.321	0.603	2.19	0.029
sexfemale:Age_Elderly:Marriagebereaved	2.539	0.564	4.497	<0.001

* Note: Male (sexmale), Young adult (AgeYoung), three days before major holiday (holidaypre), and Married (Marriagemarried) were used as references for each variable

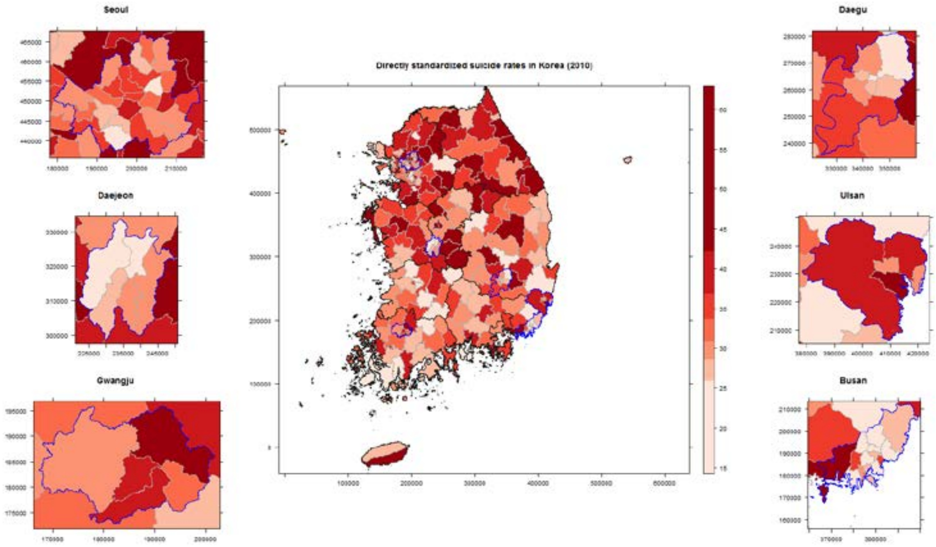
1-B. Copycat effect model

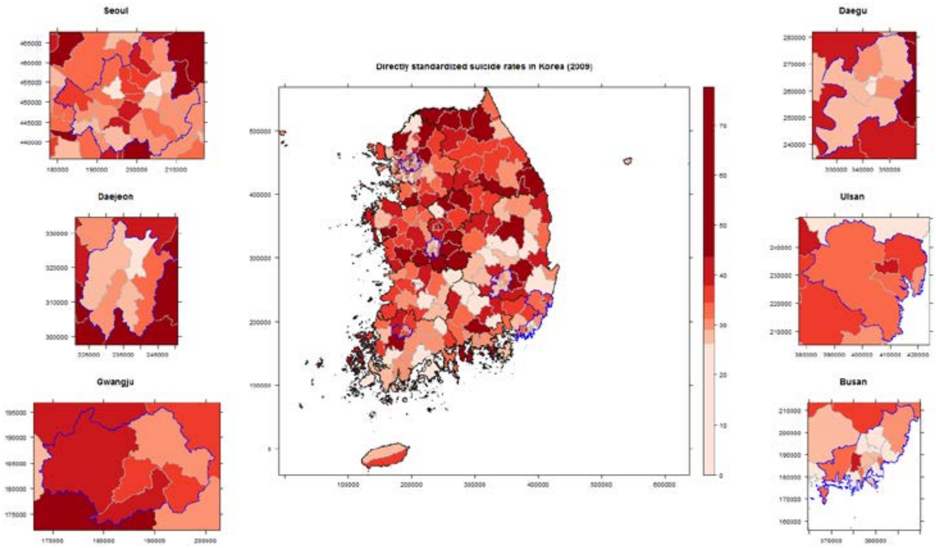
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.241	0.238	9.401	<0.001
sexfemale	−0.642	0.399	−1.608	0.108
celebpost	−0.09	0.075	−1.202	0.229
Age_YoungAdult	2.543	0.245	10.39	<0.001
Age_MiddleAged	2.627	0.244	10.764	<0.001
Age_Elderly	2.466	0.246	10.045	<0.001
methodHang	−0.677	0.361	−1.875	0.061
celebpost:methodHang	0.748	0.104	7.224	<0.001
Age_YoungAdult:methodHang	0.687	0.366	1.875	0.061
Age_MiddleAged:methodHang	0.542	0.366	1.481	0.139
Age_Elderly:methodHang	0.077	0.371	0.206	0.836
sexfemale:Age_YoungAdult	0.029	0.409	0.071	0.943
sexfemale:Age_MiddleAged	−0.248	0.411	−0.602	0.547
sexfemale:Age_Elderly	0.3	0.408	0.735	0.462
sexfemale:celebpost	0.11	0.124	0.885	0.376
sexfemale:methodHang	0.276	0.549	0.502	0.616
sexfemale:Age_YoungAdult:methodHang	0.039	0.554	0.07	0.944
sexfemale:Age_MiddleAged:methodHang	−0.834	0.56	−1.49	0.136
sexfemale:Age_Elderly:methodHang	−0.981	0.562	−1.745	0.081
sexfemale:celebpost:methodHang	0.575	0.172	3.351	0.001

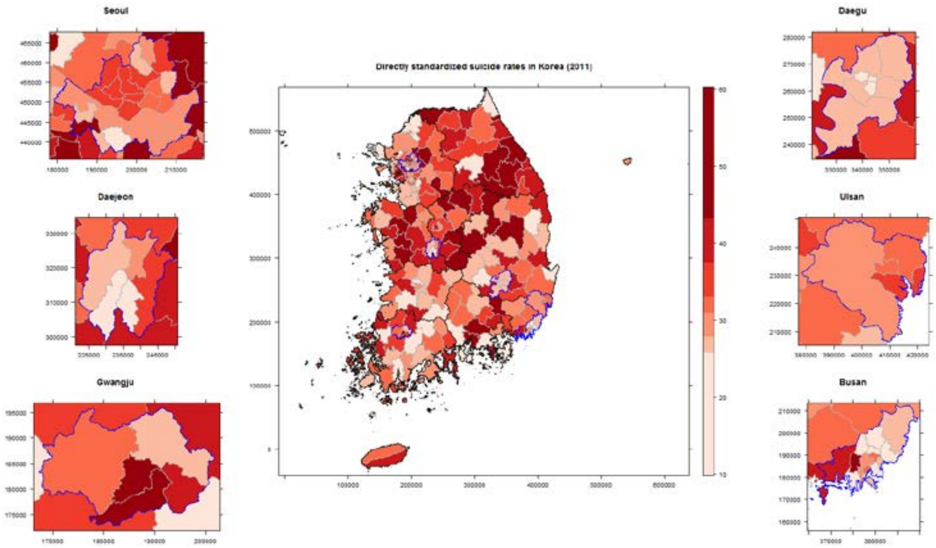
* Note: Male (sexmale), Young (Age_Young), 28-days before celebrity's suicide (celebpre), suicide methods except hanging (methodOthers) were used as references for each variable

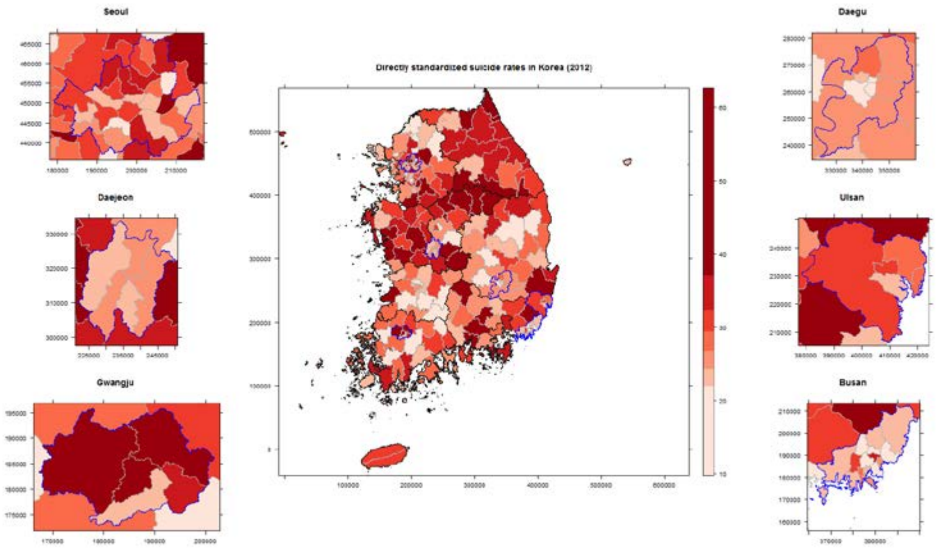
Appendix 2. Atlas of suicide in Korea

2-A. Direct standardized rates of suicide (2009-2012)

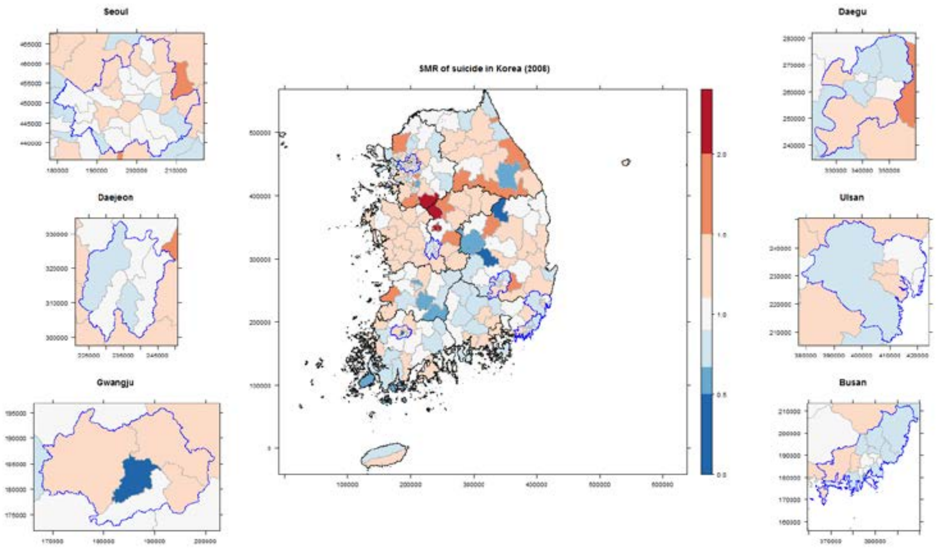


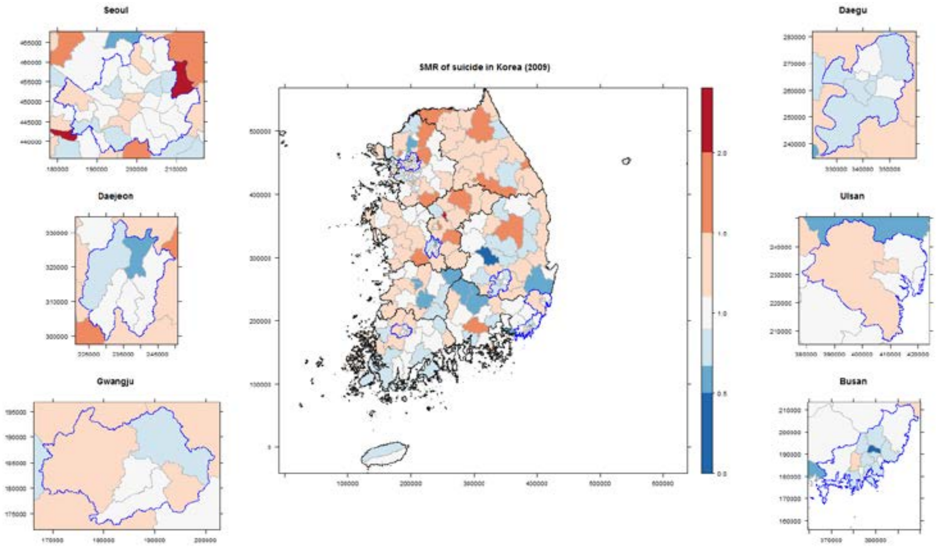


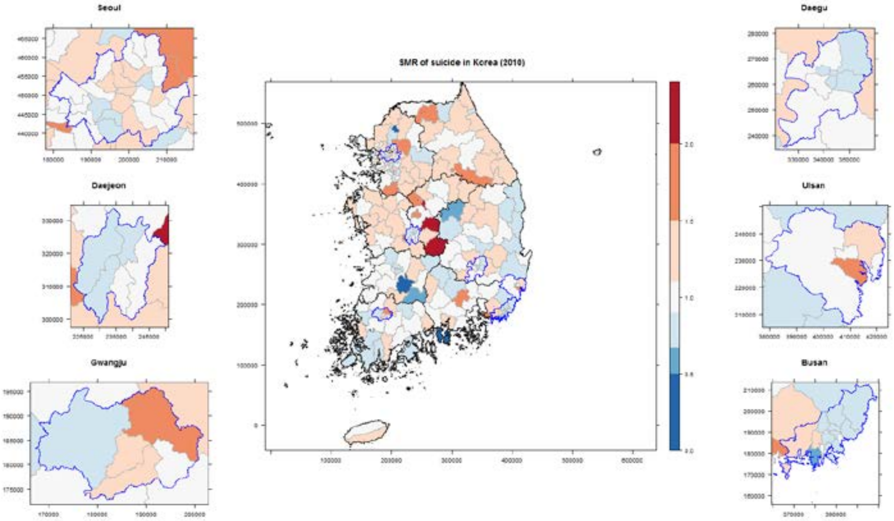

Seoul
Directly standardized suicide rates in Korea (2011)
Daegu
Daejeon
Ulsan
Gwangju
Busan

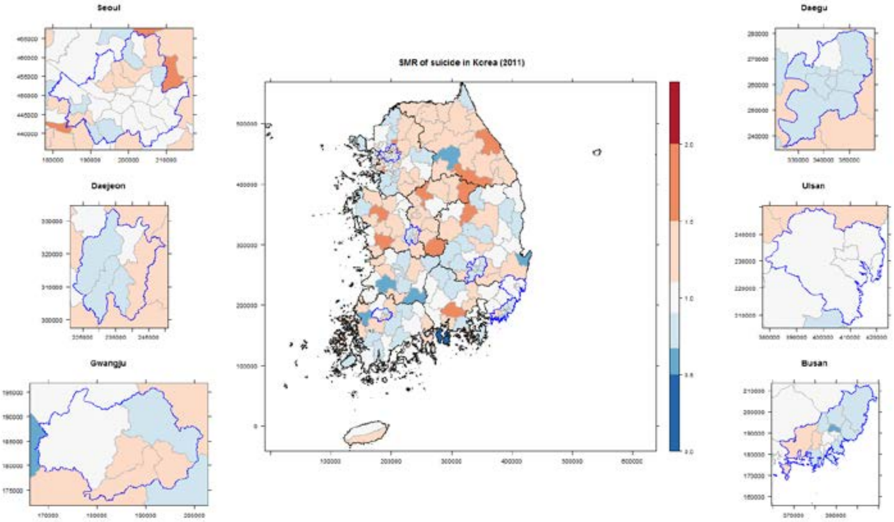

Seoul
Directly standardized suicide rates in Korea (2012)
Daegu
Daejeon
Ulsan
Gwangju
Busan

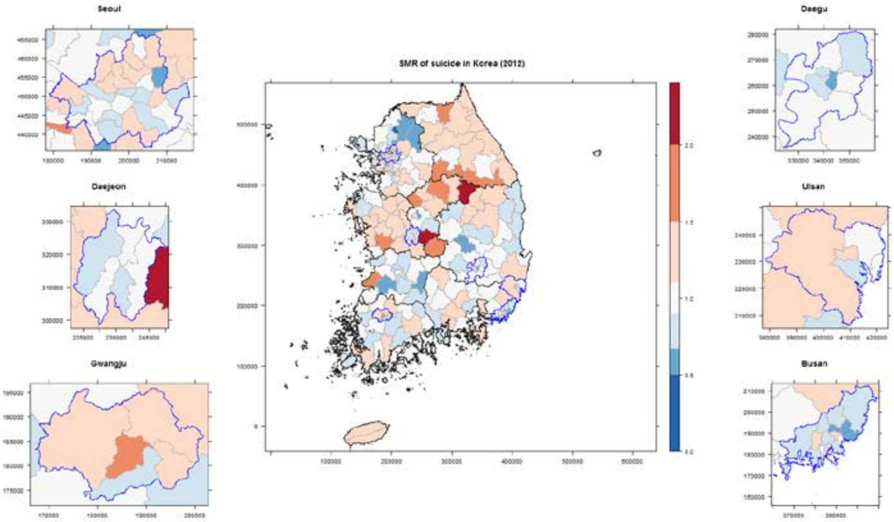
2-B. SMRs of suicide (2008-2012)



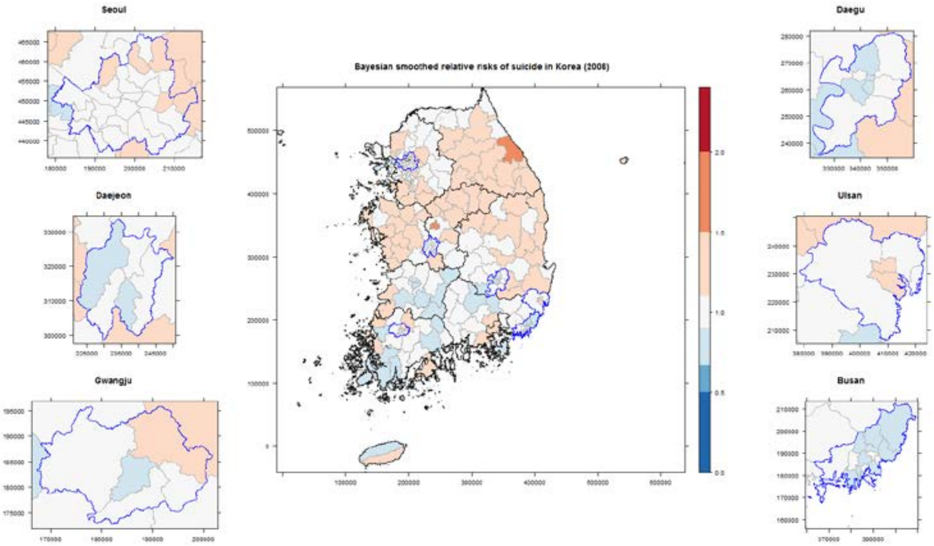


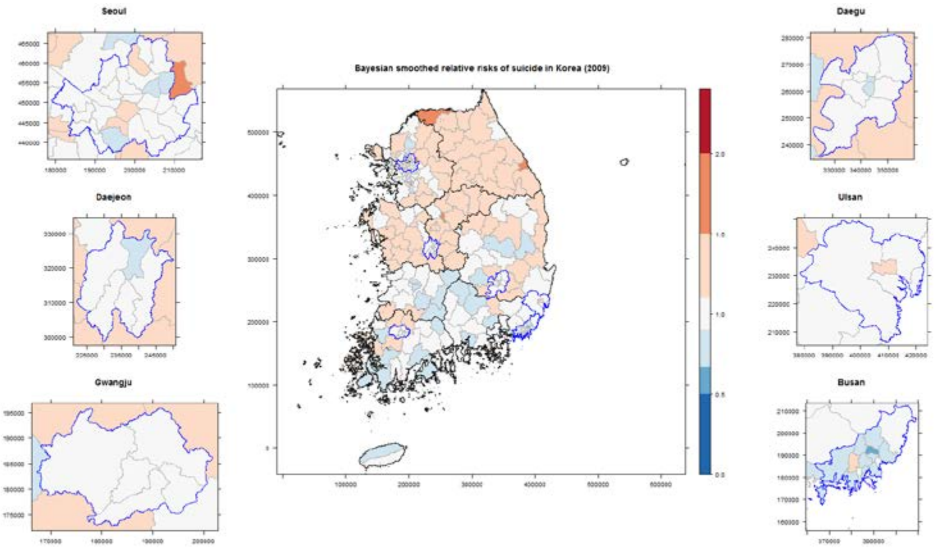

Seoul
Daegu
SMR of suicide in Korea (2010)
Daejeon
Ulsan
Gwangju
Busan

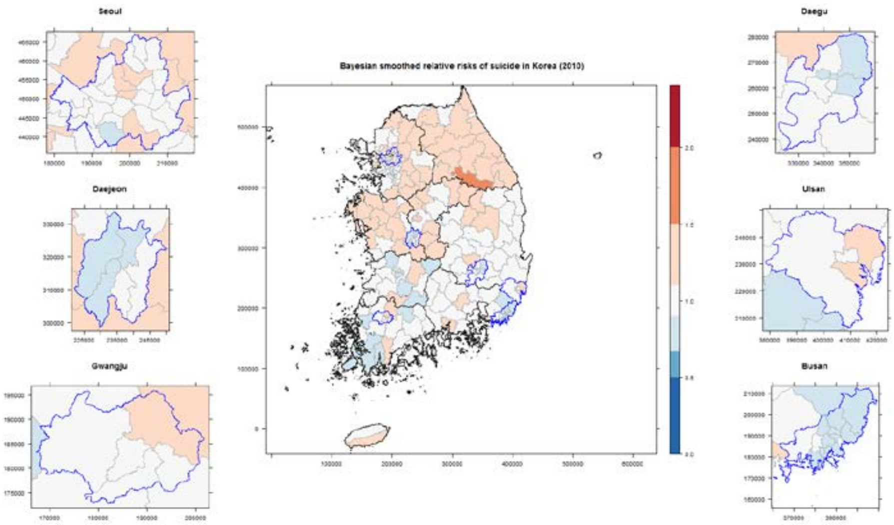

Seoul
Daegu
SMR of suicide in Korea (2011)
Daejeon
Ulsan
Gwangju
Busan

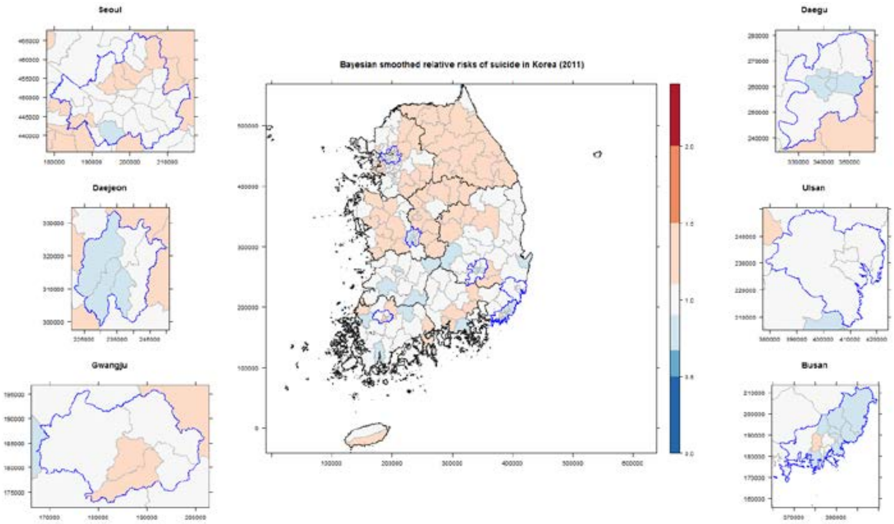

Seoul
Daegu
SMR of suicide in Korea (2012)
Daejeon
Ulsan
Gwangju
Busan

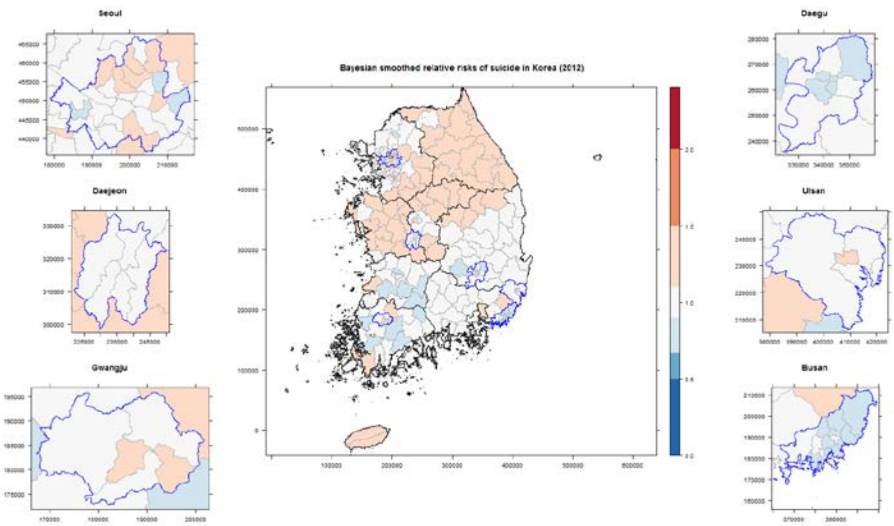
2-C. Relative risks estimated by hierarchical Bayesian conditional autoregressive model (2008-2012)



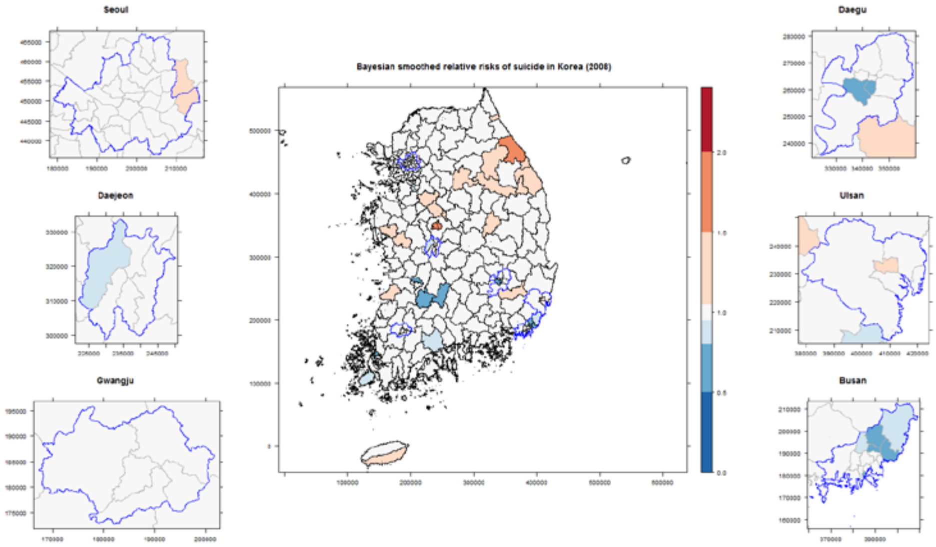


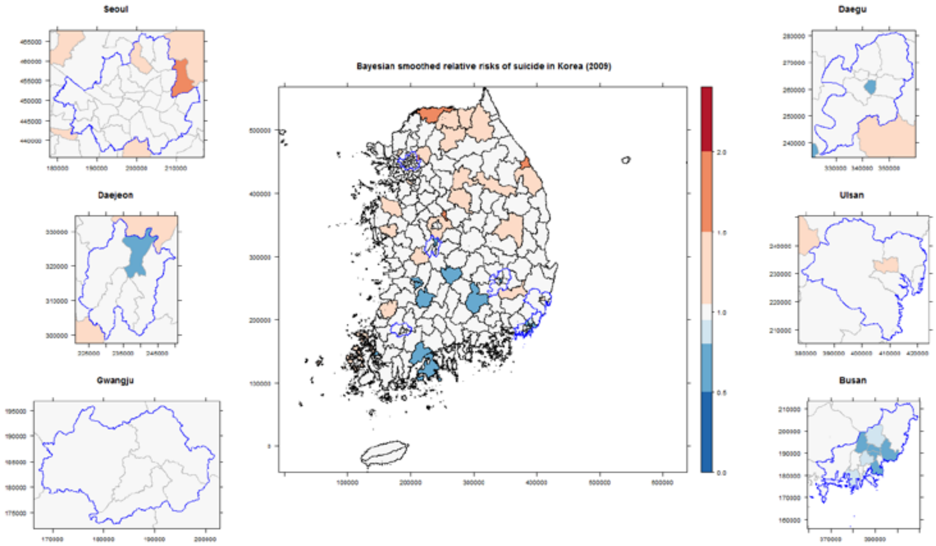

Seoul
Daejeon
Gwangju
Bayesian smoothed relative risks of suicide in Korea (2010)
Daegu
Ulsan
Busan

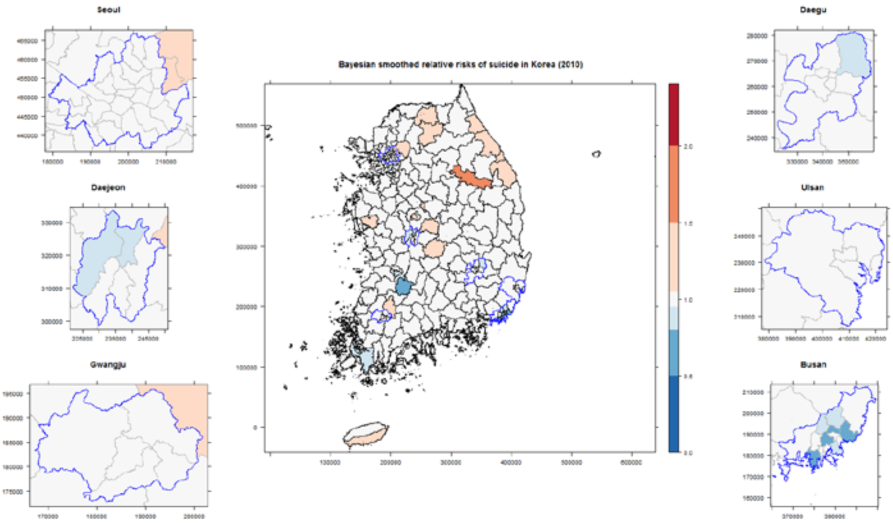

Seoul
Daejeon
Gwangju
Bayesian smoothed relative risks of suicide in Korea (2011)
Daegu
Ulsan
Busan

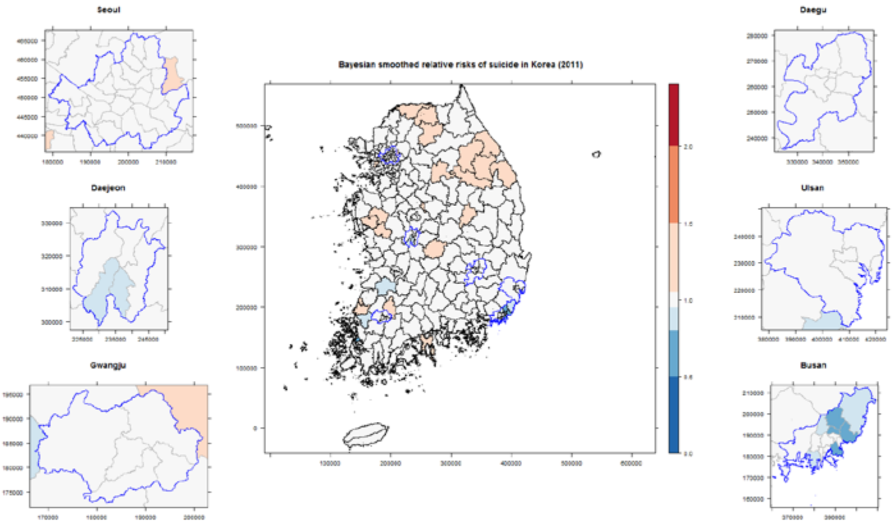

Seoul
Daejeon
Gwangju
Bayesian smoothed relative risks of suicide in Korea (2012)
Daegu
Ulsan
Busan

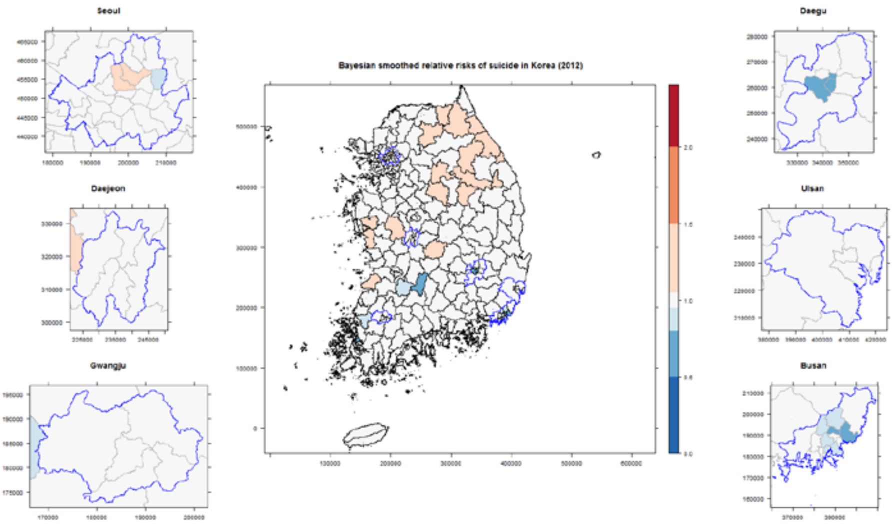
2-D. Statistically significant RRs estimated by hierarchical Bayesian conditional autoregressive model (2008-2012)






Seoul
Daejeon
Gwangju
Bayesian smoothed relative risks of suicide in Korea (2010)
Daegu
Ulsan
Busan


Seoul
Daejeon
Gwangju
Bayesian smoothed relative risks of suicide in Korea (2011)
Daegu
Ulsan
Busan


Seoul
Daejeon
Gwangju
Bayesian smoothed relative risks of suicide in Korea (2012)
Daegu
Ulsan
Busan

국 문 초 록

The Geographical Distribution and Regional Disparity of Suicide Risks in Korea, 2008-2012

Eum Young seob
Department of Geography
The Graduate School
Seoul National University

자살 문제는 국가 차원의 심각한 보건 문제이다. 본 논문은 지역 스케일에서 자살 위험의 지리적 분포를 탐색하고, 지역 간 불균등성을 측정하는 것을 목적으로 한다.

본 논문은 완료된 자살을 분석 대상으로 하며, 국가 사망원인통계를 이용하였다. 탐색적 분석을 통해 연령과 성별에 따른 자살률의 차이를 확인하는 한편, 로그선형모형을 이용하여 범주형 변수들 사이의 관계를 규명함으로써 폭력적 자살 수단의 선택, 명절 효과, 모방자살 효과와 인구사회학적 요인들 사이의 연관성을 확인하였다.

자살위험의 지리적 분포를 확인하기 위해 직접표준화법을 이용하여 시공간적으로 비교가능한 자살위험을 추정하였고, 베이지안 모형으로 통계적으로 강건한 상대위험도를 도출하였다. 범례 구성, 통계적 유의성, 카토그램 등의 지도화 방법들이 적용되었고, 구글맵과 구글어스를 인터페이스로 하는 동적 지도들이 산출되었다. 대도시권에서는 “Bull’s eye” 패턴이 발견되었고, 전국적으로는 남부 지역의 낮은 자살위험과 중부지역 이상에서의 높은 자살위험이 뚜렷하게 구분되었는데, 통계적 유의성을 고려했을 때는 강원, 충청 지역에서 높은 상대위험도를 갖는 지역들이 많이 나타났다.

지역 간 자살위험의 불균등성은 비공간적 불균등성과 공간적 군집성의 두 가지 차원을 통합한 공간비유사성지수를 이용하여 측정하고자 하였다. 먼저 각각의 차원에서 비공간적 불균등지표를

이용하여 지역 불균등성을 측정하고, 공간클러스터분석 방법을 이용하여 전역적/국지적 공간자기상관을 탐색하였다. 2008년에서 2010년까지 지역불균등성이 감소하다가 2012년까지 다시 증가하는 형태의 시계열적 변화가 나타나고, 강한 공간자기상관이 관측되었다. 이를 근거로 가우시안 커널을 이용하여 전역적 공간비유사성지수를 계산한 결과, 비공간적 불균등성의 차원에서 측정된 시계열적 변화양상과 차이가 나타나지 않았는데, 이는 자살 위험의 안정적인 시공간적 분포에 의한 것으로 볼 수 있다. 또한 행정구역을 고려하지 않고 커널 평활화 기법을 이용하여 불균등성을 측정했을 때는 시계열패턴의 변화가 나타났는데, 이는 MAUP의 존재가능성을 시사한다. 마지막으로, 지역불균등성을 자살위험지도에 효율적으로 반영하기 위한 방안으로 누적도수범례, 3D 지도화방법이 적용되었다.

주요어 : 자살위험, 상대위험도 추정, 자살지도, 지역불균등성, 공간클러스터, 공간비유사성

학 번: 2009-20152